Linear Algebra I

Matrices and Row operations

Kenneth Kuttler

Kenneth Kuttler

Linear Algebra I Matrices and Row operations

Linear Algebra I Matrices and Row operations

ISBN 978-87-403-0010-9

Contents

Preface

This is a book on linear algebra and matrix theory. While it is self contained, it will work best for those who have already had some exposure to linear algebra. It is also assumed that the reader has had calculus. Some optional topics require more analysis than this, however.

I think that the subject of linear algebra is likely the most significant topic discussed in undergraduate mathematics courses. Part of the reason for this is its usefulness in unifying so many different topics. Linear algebra is essential in analysis, applied math, and even in theoretical mathematics. This is the point of view of this book, more than a presentation of linear algebra for its own sake. This is why there are numerous applications, some fairly unusual.

This book features an ugly, elementary, and complete treatment of determinants early in the book. Thus it might be considered as Linear algebra done wrong. I have done this because of the usefulness of determinants. However, all major topics are also presented in an alternative manner which is independent of determinants.

The book has an introduction to various numerical methods used in linear algebra. This is done because of the interesting nature of these methods. The presentation here emphasizes the reasons why they work. It does not discuss many important numerical considerations necessary to use the methods effectively. These considerations are found in numerical analysis texts.

In the exercises, you may occasionally see $\uparrow$ at the beginning. This means you ought to have a look at the exercise above it. Some exercises develop a topic sequentially. There are also a few exercises which appear more than once in the book. I have done this deliberately because I think that these illustrate exceptionally important topics and because some people don't read the whole book from start to finish but instead jump in to the middle somewhere. There is one on a theorem of Sylvester which appears no fewer than 3 times. Then it is also proved in the text. There are multiple proofs of the Cayley Hamilton theorem, some in the exercises. Some exercises also are included for the sake of emphasizing something which has been done in the preceding chapter.

Preliminaries

1.1 Sets And Set Notation

A set is just a collection of things called elements. For example $\{1, 2, 3, 8\}$ would be a set consisting of the elements 1,2,3, and 8. To indicate that 3 is an element of $\{1, 2, 3, 8\}$, it is customary to write $3 \in \{1, 2, 3, 8\}$. $9 \notin \{1, 2, 3, 8\}$ means 9 is not an element of $\{1, 2, 3, 8\}$. Sometimes a rule specifies a set. For example you could specify a set as all integers larger than 2. This would be written as $S = \{x \in \mathbb{Z} : x > 2\}$. This notation says: the set of all integers, x, such that $x > 2$.

If A and B are sets with the property that every element of A is an element of B, then A is a subset of B. For example, $\{1, 2, 3, 8\}$ is a subset of $\{1, 2, 3, 4, 5, 8\}$, in symbols, $\{1, 2, 3, 8\} \subseteq \{1, 2, 3, 4, 5, 8\}$. It is sometimes said that "A is contained in B" or even "B contains A". The same statement about the two sets may also be written as $\{1, 2, 3, 4, 5, 8\} \supseteq \{1, 2, 3, 8\}$.

The union of two sets is the set consisting of everything which is an element of at least one of the sets, A or B. As an example of the union of two sets $\{1, 2, 3, 8\} \cup \{3, 4, 7, 8\} = \{1, 2, 3, 4, 7, 8\}$ because these numbers are those which are in at least one of the two sets. In general

$$A \cup B \equiv \{x : x \in A \text{ or } x \in B\}.$$

Be sure you understand that something which is in both A and B is in the union. It is not an exclusive or.

The intersection of two sets, A and B consists of everything which is in both of the sets. Thus $\{1, 2, 3, 8\} \cap \{3, 4, 7, 8\} = \{3, 8\}$ because 3 and 8 are those elements the two sets have in common. In general,

$$A \cap B \equiv \{x : x \in A \text{ and } x \in B\}.$$

The symbol $[a, b]$ where a and b are real numbers, denotes the set of real numbers x, such that $a \leq x \leq b$ and $[a, b)$ denotes the set of real numbers such that $a \leq x < b$. (a, b) consists of the set of real numbers x such that $a < x < b$ and $(a, b]$ indicates the set of numbers x such that $a < x \leq b$. $[a, \infty)$ means the set of all numbers x such that $x \geq a$ and $(-\infty, a]$ means the set of all real numbers which are less than or equal to a. These sorts of sets of real numbers are called intervals. The two points a and b are called endpoints of the interval. Other intervals such as $(-\infty, b)$ are defined by analogy to what was just explained. In general, the curved parenthesis indicates the end point it sits next to is not included while the square parenthesis indicates this end point is included. The reason that there will always be a curved parenthesis next to ∞ or $-\infty$ is that these are not real numbers. Therefore, they cannot be included in any set of real numbers.

A special set which needs to be given a name is the empty set also called the null set, denoted by $\emptyset$. Thus $\emptyset$ is defined as the set which has no elements in it. Mathematicians like to say the empty set is a subset of every set. The reason they say this is that if it were not so, there would have to exist a set A, such that $\emptyset$ has something in it which is not in A. However, $\emptyset$ has nothing in it and so the least intellectual discomfort is achieved by saying

$\emptyset \subseteq A$.

If A and B are two sets, $A \setminus B$ denotes the set of things which are in A but not in B. Thus

$$A \setminus B \equiv \{x \in A : x \notin B\}.$$

Set notation is used whenever convenient.

1.2 Functions

The concept of a function is that of something which gives a unique output for a given input.

Definition 1.2.1 *Consider two sets, D and R along with a rule which assigns a unique element of R to every element of D. This rule is called a* ***function*** *and it is denoted by a letter such as f. Given $x \in D$, $f(x)$ is the name of the thing in R which results from doing f to x. Then D is called the* ***domain*** *of f. In order to specify that D pertains to f, the* *notation $D(f)$ may be used. The set R is sometimes called the* ***range*** *of f. These days it is referred to as the* ***codomain****. The set of all elements of R which are of the form $f(x)$ for some $x \in D$ is therefore, a subset of R. This is sometimes referred to as the image of f. When this set equals R, the function f is said to be* ***onto****, also* ***surjective****. If whenever $x \neq y$ it follows $f(x) \neq f(y)$, the function is called* ***one to one.*** *, also* ***injective*** *It is common notation to write $f : D \mapsto R$ to denote the situation just described in this definition where f is a function defined on a domain D which has values in a codomain R. Sometimes you may also see something like $D \overset{f}{\mapsto} R$ to denote the same thing.*

1.3 The Number Line And Algebra Of The Real Numbers

Next, consider the real numbers, denoted by $\mathbb{R}$, as a line extending infinitely far in both directions. In this book, the notation, $\equiv$ indicates something is being defined. Thus the integers are defined as

$$\mathbb{Z} \equiv \{\cdots - 1, 0, 1, \cdots\},$$

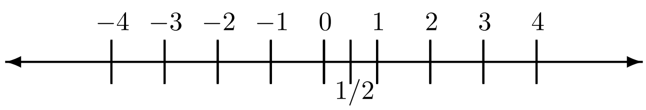
the natural numbers,

$$\mathbb{N} \equiv \{1, 2, \cdots\}$$

and the rational numbers, defined as the numbers which are the quotient of two integers.

$$\mathbb{Q} \equiv \left\{\frac{m}{n} \text{ such that } m, n \in \mathbb{Z}, n \neq 0\right\}$$

are each subsets of $\mathbb{R}$ as indicated in the following picture.



As shown in the picture, $\frac{1}{2}$ is half way between the number 0 and the number, 1. By analogy, you can see where to place all the other rational numbers. It is assumed that $\mathbb{R}$ has the following algebra properties, listed here as a collection of assertions called axioms. These properties will not be proved which is why they are called axioms rather than theorems. In general, axioms are statements which are regarded as true. Often these are things which are "self evident" either from experience or from some sort of intuition but this does not have to be the case.

Axiom 1.3.1 $x + y = y + x$, *(commutative law for addition)*

Axiom 1.3.2 $x + 0 = x$, *(additive identity).*

Axiom 1.3.3 *For each $x \in \mathbb{R}$, there exists $-x \in \mathbb{R}$ such that $x + (-x) = 0$, (existence of additive inverse).*

Axiom 1.3.4 $(x + y) + z = x + (y + z)$, *(associative law for addition).*

Axiom 1.3.5 $xy = yx$, *(commutative law for multiplication).*

Axiom 1.3.6 $(xy)z = x(yz)$, *(associative law for multiplication).*

Axiom 1.3.7 $1x = x$, *(multiplicative identity).*

Axiom 1.3.8 *For each* $x \neq 0$, *there exists* x^{-1} *such that* $xx^{-1} = 1$.*(existence of multiplicative inverse).*

Axiom 1.3.9 $x(y+z) = xy + xz$.*(distributive law).*

These axioms are known as the field axioms and any set (there are many others besides $\mathbb{R}$) which has two such operations satisfying the above axioms is called a field. Division and subtraction are defined in the usual way by $x - y \equiv x + (-y)$ and $x/y \equiv x\left(y^{-1}\right)$.

Here is a little proposition which derives some familiar facts.

Proposition 1.3.10 0 *and* 1 *are unique. Also* $-x$ *is unique and* x^{-1} *is unique. Furthermore,* $0x = x0 = 0$ *and* $-x = (-1)x$.

Proof: Suppose $0'$ is another additive identity. Then

$$0' = 0' + 0 = 0.$$

Thus 0 is unique. Say $1'$ is another multiplicative identity. Then

$$1 = 1'1 = 1'.$$

Now suppose y acts like the additive inverse of x. Then

$$-x = (-x) + 0 = (-x) + (x+y) = (-x+x) + y = y$$

Finally,

$$0x = (0+0)x = 0x + 0x$$

and so

$$0 = -(0x) + 0x = -(0x) + (0x + 0x) = (-(0x) + 0x) + 0x = 0x$$

Finally

$$x + (-1)x = (1 + (-1))x = 0x = 0$$

and so by uniqueness of the additive inverse, $(-1)x = -x$. ■

1.4 Ordered fields

The real numbers $\mathbb{R}$ are an example of an ordered field. More generally, here is a definition.

Definition 1.4.1 *Let* F *be a field. It is an ordered field if there exists an order,* $<$ *which satisfies*

1. *For any* $x \neq y$, *either* $x < y$ *or* $y < x$.
2. *If* $x < y$ *and either* $z < w$ *or* $z = w$, *then,* $x + z < y + w$.
3. *If* $0 < x, 0 < y$, *then* $xy > 0$.

With this definition, the familiar properties of order can be proved. The following proposition lists many of these familiar properties. The relation '$a > b$' has the same meaning as '$b < a$'.

Proposition 1.4.2 *The following are obtained.*

1. *If* $x < y$ *and* $y < z$, *then* $x < z$.
2. *If* $x > 0$ *and* $y > 0$, *then* $x + y > 0$.
3. *If* $x > 0$, *then* $-x < 0$.
4. *If* $x \neq 0$, *either* x *or* $-x$ *is* > 0.
5. *If* $x < y$, *then* $-x > -y$.
6. *If* $x \neq 0$, *then* $x^2 > 0$.
7. *If* $0 < x < y$ *then* $x^{-1} > y^{-1}$.

Proof: First consider 1, called the transitive law. Suppose that $x < y$ and $y < z$. Then from the axioms, $x + y < y + z$ and so, adding $-y$ to both sides, it follows

$$x < z$$

Next consider 2. Suppose $x > 0$ and $y > 0$. Then from 2,

$$0 = 0 + 0 < x + y.$$

Next consider 3. It is assumed $x > 0$ so

$$0 = -x + x > 0 + (-x) = -x$$

Now consider 4. If $x < 0$, then

$$0 = x + (-x) < 0 + (-x) = -x.$$

Consider the 5. Since $x < y$, it follows from 2

$$0 = x + (-x) < y + (-x)$$

and so by 4 and Proposition 1.3.10,

$$(-1)(y + (-x)) < 0$$

Also from Proposition 1.3.10 $(-1)(-x) = -(-x) = x$ and so

$$-y + x < 0.$$

Hence

$$-y < -x.$$

Consider 6. If $x > 0$, there is nothing to show. It follows from the definition. If $x < 0$, then by 4, $-x > 0$ and so by Proposition 1.3.10 and the definition of the order,

$$(-x)^2 = (-1)(-1)x^2 > 0$$

By this proposition again, $(-1)(-1) = -(-1) = 1$ and so $x^2 > 0$ as claimed. Note that $1 > 0$ because it equals 1^2.

Finally, consider 7. First, if $x > 0$ then if $x^{-1} < 0$, it would follow $(-1)x^{-1} > 0$ and so $x(-1)x^{-1} = (-1)1 = -1 > 0$. However, this would require

$$0 > 1 = 1^2 > 0$$

from what was just shown. Therefore, $x^{-1} > 0$. Now the assumption implies $y + (-1)x > 0$ and so multiplying by x^{-1},

$$yx^{-1} + (-1)xx^{-1} = yx^{-1} + (-1) > 0$$

Now multiply by y^{-1}, which by the above satisfies $y^{-1} > 0$, to obtain

$$x^{-1} + (-1)y^{-1} > 0$$

and so

$$x^{-1} > y^{-1}. \blacksquare$$

In an ordered field the symbols $\leq$ and $\geq$ have the usual meanings. Thus $a \leq b$ means $a < b$ or else $a = b$, etc.

1.5 The Complex Numbers

Just as a real number should be considered as a point on the line, a complex number is considered a point in the plane which can be identified in the usual way using the Cartesian coordinates of the point. Thus (a, b) identifies a point whose x coordinate is a and whose y coordinate is b. In dealing with complex numbers, such a point is written as $a + ib$ and multiplication and addition are defined in the most obvious way subject to the convention that $i^2 = -1$. Thus,

$$(a + ib) + (c + id) = (a + c) + i(b + d)$$

and

$$\begin{aligned}(a + ib)(c + id) &= ac + iad + ibc + i^2bd \\ &= (ac - bd) + i(bc + ad).\end{aligned}$$

Every non zero complex number, $a+ib$, with $a^2+b^2 \neq 0$, has a unique multiplicative inverse.

$$\frac{1}{a + ib} = \frac{a - ib}{a^2 + b^2} = \frac{a}{a^2 + b^2} - i\frac{b}{a^2 + b^2}.$$

You should prove the following theorem.

Theorem 1.5.1 *The complex numbers with multiplication and addition defined as above form a field satisfying all the field axioms listed on Page 13.*

Note that if $x + iy$ is a complex number, it can be written as

$$x + iy = \sqrt{x^2 + y^2}\left(\frac{x}{\sqrt{x^2 + y^2}} + i\frac{y}{\sqrt{x^2 + y^2}}\right)$$

Now $\left(\frac{x}{\sqrt{x^2+y^2}}, \frac{y}{\sqrt{x^2+y^2}}\right)$ is a point on the unit circle and so there exists a unique $\theta \in [0, 2\pi)$ such that this ordered pair equals $(\cos\theta, \sin\theta)$. Letting $r = \sqrt{x^2+y^2}$, it follows that the complex number can be written in the form

$$x + iy = r(\cos\theta + i\sin\theta)$$

This is called the polar form of the complex number.

The field of complex numbers is denoted as $\mathbb{C}$. An important construction regarding complex numbers is the complex conjugate denoted by a horizontal line above the number. It is defined as follows.

$$\overline{a+ib} \equiv a - ib.$$

What it does is reflect a given complex number across the x axis. Algebraically, the following formula is easy to obtain.

$$\left(\overline{a+ib}\right)(a+ib) = a^2 + b^2.$$

Definition 1.5.2 *Define the absolute value of a complex number as follows.*

$$|a+ib| \equiv \sqrt{a^2+b^2}.$$

Thus, denoting by z the complex number, $z = a + ib$,

$$|z| = (z\overline{z})^{1/2}.$$

With this definition, it is important to note the following. Be sure to verify this. It is not too hard but you need to do it.

Remark 1.5.3 *: Let $z = a+ib$ and $w = c+id$. Then $|z-w| = \sqrt{(a-c)^2 + (b-d)^2}$. Thus the distance between the point in the plane determined by the ordered pair, (a, b) and the ordered pair (c, d) equals $|z-w|$ where z and w are as just described.*

For example, consider the distance between $(2, 5)$ and $(1, 8)$. From the distance formula this distance equals $\sqrt{(2-1)^2 + (5-8)^2} = \sqrt{10}$. On the other hand, letting $z = 2+i5$ and $w = 1+i8$, $z-w = 1-i3$ and so $(z-w)\left(\overline{z-w}\right) = (1-i3)(1+i3) = 10$ so $|z-w| = \sqrt{10}$, the same thing obtained with the distance formula.

Complex numbers, are often written in the so called polar form which is described next. Suppose $x + iy$ is a complex number. Then

$$x + iy = \sqrt{x^2+y^2}\left(\frac{x}{\sqrt{x^2+y^2}} + i\frac{y}{\sqrt{x^2+y^2}}\right).$$

Now note that

$$\left(\frac{x}{\sqrt{x^2+y^2}}\right)^2 + \left(\frac{y}{\sqrt{x^2+y^2}}\right)^2 = 1$$

and so

$$\left(\frac{x}{\sqrt{x^2+y^2}}, \frac{y}{\sqrt{x^2+y^2}}\right)$$

is a point on the unit circle. Therefore, there exists a unique angle, $\theta \in [0, 2\pi)$ such that

$$\cos\theta = \frac{x}{\sqrt{x^2+y^2}}, \sin\theta = \frac{y}{\sqrt{x^2+y^2}}.$$

The polar form of the complex number is then

$$r(\cos\theta + i\sin\theta)$$

where θ is this angle just described and $r = \sqrt{x^2+y^2}$.

A fundamental identity is the formula of De Moivre which follows.

Theorem 1.5.4 *Let $r > 0$ be given. Then if n is a positive integer,*

$$[r(\cos t + i\sin t)]^n = r^n(\cos nt + i\sin nt).$$

Proof: It is clear the formula holds if $n = 1$. Suppose it is true for n.

$$[r(\cos t + i\sin t)]^{n+1} = [r(\cos t + i\sin t)]^n [r(\cos t + i\sin t)]$$

which by induction equals

$$= r^{n+1}(\cos nt + i\sin nt)(\cos t + i\sin t)$$

$$= r^{n+1}((\cos nt\cos t - \sin nt\sin t) + i(\sin nt\cos t + \cos nt\sin t))$$

$$= r^{n+1}(\cos(n+1)t + i\sin(n+1)t)$$

by the formulas for the cosine and sine of the sum of two angles. ■

Corollary 1.5.5 *Let z be a non zero complex number. Then there are always exactly k k^{th} roots of z in $\mathbb{C}$.*

Proof: Let $z = x + iy$ and let $z = |z|(\cos t + i\sin t)$ be the polar form of the complex number. By De Moivre's theorem, a complex number,

$$r(\cos\alpha + i\sin\alpha),$$

is a k^{th} root of z if and only if

$$r^k(\cos k\alpha + i\sin k\alpha) = |z|(\cos t + i\sin t).$$

This requires $r^k = |z|$ and so $r = |z|^{1/k}$ and also both $\cos(k\alpha) = \cos t$ and $\sin(k\alpha) = \sin t$. This can only happen if

$$k\alpha = t + 2l\pi$$

for l an integer. Thus

$$\alpha = \frac{t+2l\pi}{k}, l \in \mathbb{Z}$$

and so the k^{th} roots of z are of the form

$$|z|^{1/k}\left(\cos\left(\frac{t+2l\pi}{k}\right) + i\sin\left(\frac{t+2l\pi}{k}\right)\right),\ l \in \mathbb{Z}.$$

Since the cosine and sine are periodic of period 2π, there are exactly k distinct numbers

$\emptyset \subseteq A$.

If A and B are two sets, $A \setminus B$ denotes the set of things which are in A but not in B. Thus

$$A \setminus B \equiv \{x \in A : x \notin B\} .$$

Set notation is used whenever convenient.

1.2 Functions

The concept of a function is that of something which gives a unique output for a given input.

Definition 1.2.1 *Consider two sets, D and R along with a rule which assigns a unique element of R to every element of D. This rule is called a* ***function*** *and it is denoted by a letter such as f. Given $x \in D$, $f(x)$ is the name of the thing in R which results from doing f to x. Then D is called the* ***domain*** *of f. In order to specify that D pertains to f, the*

Thus the cube roots of i are $\frac{\sqrt{3}}{2}+i\left(\frac{1}{2}\right)$, $\frac{-\sqrt{3}}{2}+i\left(\frac{1}{2}\right)$, and $-i$.

The ability to find k^{th} roots can also be used to factor some polynomials.

Example 1.5.7 *Factor the polynomial* x^3-27.

First find the cube roots of 27. By the above procedure using De Moivre's theorem, these cube roots are $3, 3\left(\frac{-1}{2}+i\frac{\sqrt{3}}{2}\right)$, and $3\left(\frac{-1}{2}-i\frac{\sqrt{3}}{2}\right)$. Therefore, $x^3+27=$

$$(x-3)\left(x-3\left(\frac{-1}{2}+i\frac{\sqrt{3}}{2}\right)\right)\left(x-3\left(\frac{-1}{2}-i\frac{\sqrt{3}}{2}\right)\right).$$

Note also $\left(x-3\left(\frac{-1}{2}+i\frac{\sqrt{3}}{2}\right)\right)\left(x-3\left(\frac{-1}{2}-i\frac{\sqrt{3}}{2}\right)\right)=x^2+3x+9$ and so

$$x^3-27=(x-3)\left(x^2+3x+9\right)$$

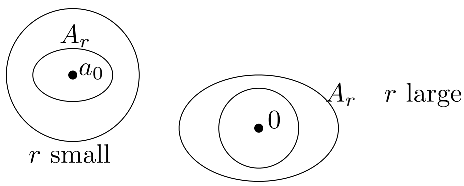
where the quadratic polynomial, x^2+3x+9 cannot be factored without using complex numbers.

The real and complex numbers both are fields satisfying the axioms on Page 13 and it is usually one of these two fields which is used in linear algebra. The numbers are often called scalars. However, it turns out that all algebraic notions work for any field and there are many others. For this reason, I will often refer to the field of scalars as $\mathbb{F}$ although $\mathbb{F}$ will usually be either the real or complex numbers. If there is any doubt, assume it is the field of complex numbers which is meant. The reason the complex numbers are so significant in linear algebra is that they are algebraically complete. This means that every polynomial $\sum_{k=0}^{n} a_k z^k,\ n\geq 1, a_n\neq 0$, having coefficients a_k in $\mathbb{C}$ has a root in in $\mathbb{C}$.

Later in the book, proofs of the fundamental theorem of algebra are given. However, here is a simple explanation of why you should believe this theorem. The issue is whether there exists $z\in\mathbb{C}$ such that $p(z)=0$ for $p(z)$ a polynomial having coefficients in $\mathbb{C}$. Dividing by the leading coefficient, we can assume that $p(z)$ is of the form

$$p(z)=z^n+a_{n-1}z^{n-1}+\cdots+a_1z+a_0,\ a_0\neq 0.$$

If $a_0=0$, there is nothing to prove. Denote by C_r the circle of radius r in the complex plane which is centered at 0. Then if r is sufficiently large and $|z|=r$, the term z^n is far larger than the rest of the polynomial. Thus, for r large enough, $A_r=\{p(z):z\in C_r\}$ describes a closed curve which misses the inside of some circle having 0 as its center. Now shrink r. Eventually, for r small enough, the non constant terms are negligible and so A_r is a curve which is contained in some circle centered at a_0 which has 0 in its outside.



Thus it is reasonable to believe that for some r during this shrinking process, the set A_r must hit 0. It follows that $p(z)=0$ for some z. This is one of those arguments which seems all right until you think about it too much. Nevertheless, it will suffice to see that the fundamental theorem of algebra is at least very plausible. A complete proof is in an appendix.

1.6 Exercises

1. Let $z = 5 + i9$. Find z^{-1}.

2. Let $z = 2 + i7$ and let $w = 3 - i8$. Find $zw, z + w, z^2$, and w/z.

3. Give the complete solution to $x^4 + 16 = 0$.

4. Graph the complex cube roots of 8 in the complex plane. Do the same for the four fourth roots of 16.

5. If z is a complex number, show there exists ω a complex number with $|\omega| = 1$ and $\omega z = |z|$.

6. De Moivre's theorem says $[r(\cos t + i \sin t)]^n = r^n (\cos nt + i \sin nt)$ for n a positive integer. Does this formula continue to hold for all integers, n, even negative integers? Explain.

7. You already know formulas for $\cos(x + y)$ and $\sin(x + y)$ and these were used to prove De Moivre's theorem. Now using De Moivre's theorem, derive a formula for $\sin(5x)$ and one for $\cos(5x)$. **Hint:** Use the binomial theorem.

8. If z and w are two complex numbers and the polar form of z involves the angle θ while the polar form of w involves the angle ϕ, show that in the polar form for zw the angle involved is $\theta + \phi$. Also, show that in the polar form of a complex number, z, $r = |z|$.

9. Factor $x^3 + 8$ as a product of linear factors.

10. Write $x^3 + 27$ in the form $(x + 3)\left(x^2 + ax + b\right)$ where $x^2 + ax + b$ cannot be factored any more using only real numbers.

11. Completely factor $x^4 + 16$ as a product of linear factors.

12. Factor $x^4 + 16$ as the product of two quadratic polynomials each of which cannot be factored further without using complex numbers.

13. If z, w are complex numbers prove $\overline{zw} = \overline{z}\,\overline{w}$ and then show by induction that $\overline{z_1 \cdots z_m} = \overline{z_1} \cdots \overline{z_m}$. Also verify that $\overline{\sum_{k=1}^{m} z_k} = \sum_{k=1}^{m} \overline{z_k}$. In words this says the conjugate of a product equals the product of the conjugates and the conjugate of a sum equals the sum of the conjugates.

14. Suppose $p(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$ where all the a_k are real numbers. Suppose also that $p(z) = 0$ for some $z \in \mathbb{C}$. Show it follows that $p(\overline{z}) = 0$ also.

15. I claim that $1 = -1$. Here is why.

$$-1 = i^2 = \sqrt{-1}\sqrt{-1} = \sqrt{(-1)^2} = \sqrt{1} = 1.$$

This is clearly a remarkable result but is there something wrong with it? If so, what is wrong?

16. De Moivre's theorem is really a grand thing. I plan to use it now for rational exponents, not just integers.

$$1 = 1^{(1/4)} = (\cos 2\pi + i \sin 2\pi)^{1/4} = \cos(\pi/2) + i \sin(\pi/2) = i.$$

Therefore, squaring both sides it follows $1 = -1$ as in the previous problem. What does this tell you about De Moivre's theorem? Is there a profound difference between raising numbers to integer powers and raising numbers to non integer powers?

17. Show that $\mathbb{C}$ cannot be considered an ordered field. **Hint:** Consider $i^2 = -1$. Recall that $1 > 0$ by Proposition 1.4.2.

18. Say $a + ib < x + iy$ if $a < x$ or if $a = x$, then $b < y$. This is called the lexicographic order. Show that any two different complex numbers can be compared with this order. What goes wrong in terms of the other requirements for an ordered field.

19. With the order of Problem 18, consider for $n \in \mathbb{N}$ the complex number $1 - \frac{1}{n}$. Show that with the lexicographic order just described, each of $1 - in$ is an upper bound to all these numbers. Therefore, this is a set which is "bounded above" but has no least upper bound with respect to the lexicographic order on $\mathbb{C}$.

1.7 Completeness of $\mathbb{R}$

Recall the following important definition from calculus, completeness of $\mathbb{R}$.

Definition 1.7.1 *A non empty set, $S \subseteq \mathbb{R}$ is bounded above (below) if there exists $x \in \mathbb{R}$ such that $x \geq (\leq) s$ for all $s \in S$. If S is a nonempty set in $\mathbb{R}$ which is bounded above, then a number, l which has the property that l is an upper bound and that every other upper bound is no smaller than l is called a least upper bound, l.u.b. (S) or often $\sup(S)$. If S is a nonempty set bounded below, define the greatest lower bound, g.l.b. (S) or $\inf(S)$ similarly. Thus g is the g.l.b. (S) means g is a lower bound for S and it is the largest of all lower bounds. If S is a nonempty subset of $\mathbb{R}$ which is not bounded above, this information is expressed by saying $\sup(S) = +\infty$ and if S is not bounded below, $\inf(S) = -\infty$.*

Every existence theorem in calculus depends on some form of the completeness axiom.

Axiom 1.7.2 *(completeness) Every nonempty set of real numbers which is bounded above has a least upper bound and every nonempty set of real numbers which is bounded below has a greatest lower bound.*

It is this axiom which distinguishes Calculus from Algebra. A fundamental result about sup and inf is the following.

Proposition 1.7.3 *Let S be a nonempty set and suppose* $\sup(S)$ *exists. Then for every* $\delta > 0$,

$$S \cap (\sup(S) - \delta, \sup(S)] \neq \emptyset.$$

If $\inf(S)$ *exists, then for every* $\delta > 0$,

$$S \cap [\inf(S), \inf(S) + \delta) \neq \emptyset.$$

Proof: Consider the first claim. If the indicated set equals $\emptyset$, then $\sup(S) - \delta$ is an upper bound for S which is smaller than $\sup(S)$, contrary to the definition of $\sup(S)$ as the least upper bound. In the second claim, if the indicated set equals $\emptyset$, then $\inf(S) + \delta$ would be a lower bound which is larger than $\inf(S)$ contrary to the definition of $\inf(S)$. ■

1.8 Well Ordering And Archimedean Property

Definition 1.8.1 *A set is well ordered if every nonempty subset S, contains a smallest element z having the property that $z \leq x$ for all $x \in S$.*

Axiom 1.8.2 *Any set of integers larger than a given number is well ordered.*

In particular, the natural numbers defined as

$$\mathbb{N} \equiv \{1, 2, \cdots\}$$

is well ordered.

The above axiom implies the principle of mathematical induction.

Theorem 1.8.3 *(Mathematical induction) A set $S \subseteq \mathbb{Z}$, having the property that $a \in S$ and $n + 1 \in S$ whenever $n \in S$ contains all integers $x \in \mathbb{Z}$ such that $x \geq a$.*

Proof: Let $T \equiv ([a, \infty) \cap \mathbb{Z}) \setminus S$. Thus T consists of all integers larger than or equal to a which are not in S. The theorem will be proved if $T = \emptyset$. If $T \neq \emptyset$ then by the well ordering principle, there would have to exist a smallest element of T, denoted as b. It must be the case that $b > a$ since by definition, $a \notin T$. Then the integer, $b - 1 \geq a$ and $b - 1 \notin S$ because if $b - 1 \in S$, then $b - 1 + 1 = b \in S$ by the assumed property of S. Therefore, $b - 1 \in ([a, \infty) \cap \mathbb{Z}) \setminus S = T$ which contradicts the choice of b as the smallest element of T. ($b - 1$ is smaller.) Since a contradiction is obtained by assuming $T \neq \emptyset$, it must be the case that $T = \emptyset$ and this says that everything in $[a, \infty) \cap \mathbb{Z}$ is also in S. ■

Example 1.8.4 *Show that for all $n \in \mathbb{N}$, $\frac{1}{2} \cdot \frac{3}{4} \cdots \frac{2n-1}{2n} < \frac{1}{\sqrt{2n+1}}$.*

If $n = 1$ this reduces to the statement that $\frac{1}{2} < \frac{1}{\sqrt{3}}$ which is obviously true. Suppose then that the inequality holds for n. Then

$$\begin{aligned} \frac{1}{2} \cdot \frac{3}{4} \cdots \frac{2n-1}{2n} \cdot \frac{2n+1}{2n+2} &< \frac{1}{\sqrt{2n+1}} \frac{2n+1}{2n+2} \\ &= \frac{\sqrt{2n+1}}{2n+2}. \end{aligned}$$

The theorem will be proved if this last expression is less than $\frac{1}{\sqrt{2n+3}}$. This happens if and only if

$$\left(\frac{1}{\sqrt{2n+3}} \right)^2 = \frac{1}{2n+3} > \frac{2n+1}{(2n+2)^2}$$

which occurs if and only if $(2n+2)^2 > (2n+3)(2n+1)$ and this is clearly true which may be seen from expanding both sides. This proves the inequality.

Definition 1.8.5 *The Archimedean property states that whenever $x \in \mathbb{R}$, and $a > 0$, there exists $n \in \mathbb{N}$ such that $na > x$.*

Proposition 1.8.6 *$\mathbb{R}$ has the Archimedean property.*

Proof: Suppose it is not true. Then there exists $x \in \mathbb{R}$ and $a > 0$ such that $na \leq x$ for all $n \in \mathbb{N}$. Let $S = \{na : n \in \mathbb{N}\}$. By assumption, this is bounded above by x. By completeness, it has a least upper bound y. By Proposition 1.7.3 there exists $n \in \mathbb{N}$ such that

$$y - a < na \leq y.$$

Then $y = y - a + a < na + a = (n+1) a \leq y$, a contradiction. ■

Theorem 1.8.7 *Suppose $x < y$ and $y - x > 1$. Then there exists an integer $l \in \mathbb{Z}$, such that $x < l < y$. If x is an integer, there is no integer y satisfying $x < y < x + 1$.*

Proof: Let x be the smallest positive integer. Not surprisingly, $x = 1$ but this can be proved. If $x < 1$ then $x^2 < x$ contradicting the assertion that x is the smallest natural number. Therefore, 1 is the smallest natural number. This shows there is no integer, y, satisfying $x < y < x + 1$ since otherwise, you could subtract x and conclude $0 < y - x < 1$ for some integer $y - x$.

Now suppose $y - x > 1$ and let

$$S \equiv \{w \in \mathbb{N} : w \geq y\}.$$

The set S is nonempty by the Archimedean property. Let k be the smallest element of S. Therefore, $k - 1 < y$. Either $k - 1 \leq x$ or $k - 1 > x$. If $k - 1 \leq x$, then

$$y - x \leq y - (k - 1) = \overbrace{y - k}^{\leq 0} + 1 \leq 1$$

contrary to the assumption that $y - x > 1$. Therefore, $x < k - 1 < y$. Let $l = k - 1$. ■

It is the next theorem which gives the density of the rational numbers. This means that for any real number, there exists a rational number arbitrarily close to it.

Theorem 1.8.8 *If $x < y$ then there exists a rational number r such that $x < r < y$.*

Proof: Let $n \in \mathbb{N}$ be large enough that

$$n(y - x) > 1.$$

Thus $(y - x)$ added to itself n times is larger than 1. Therefore,

$$n(y - x) = ny + n(-x) = ny - nx > 1.$$

It follows from Theorem 1.8.7 there exists $m \in \mathbb{Z}$ such that

$$nx < m < ny$$

and so take $r = m/n$. ■

Definition 1.8.9 *A set $S \subseteq \mathbb{R}$ is dense in $\mathbb{R}$ if whenever $a < b$, $S \cap (a, b) \neq \emptyset$.*

Thus the above theorem says $\mathbb{Q}$ is "dense" in $\mathbb{R}$.

Theorem 1.8.10 *Suppose $0 < a$ and let $b \geq 0$. Then there exists a unique integer p and real number r such that $0 \leq r < a$ and $b = pa + r$.*

Proof: Let $S \equiv \{n \in \mathbb{N} : an > b\}$. By the Archimedean property this set is nonempty. Let $p + 1$ be the smallest element of S. Then $pa \leq b$ because $p + 1$ is the smallest in S. Therefore,

$$r \equiv b - pa \geq 0.$$

If $r \geq a$ then $b - pa \geq a$ and so $b \geq (p + 1)a$ contradicting $p + 1 \in S$. Therefore, $r < a$ as desired.

To verify uniqueness of p and r, suppose p_i and r_i, $i = 1, 2$, both work and $r_2 > r_1$. Then a little algebra shows

$$p_1 - p_2 = \frac{r_2 - r_1}{a} \in (0, 1).$$

Thus $p_1 - p_2$ is an integer between 0 and 1, contradicting Theorem 1.8.7. The case that $r_1 > r_2$ cannot occur either by similar reasoning. Thus $r_1 = r_2$ and it follows that $p_1 = p_2$. ■

This theorem is called the Euclidean algorithm when a and b are integers.

1.9 Division And Numbers

First recall Theorem 1.8.10, the Euclidean algorithm.

Theorem 1.9.1 *Suppose* $0 < a$ *and let* $b \geq 0$*. Then there exists a unique integer* p *and real number* r *such that* $0 \leq r < a$ *and* $b = pa + r$*.*

The following definition describes what is meant by a prime number and also what is meant by the word "divides".

Definition 1.9.2 *The number,* a *divides the number,* b *if in Theorem 1.8.10,* $r = 0$*. That is there is zero remainder. The notation for this is* $a|b$*, read* a *divides* b *and* a *is called a factor of* b*. A prime number is one which has the property that the only numbers which divide it are itself and 1. The greatest common divisor of two positive integers,* m, n *is that number,* p *which has the property that* p *divides both* m *and* n *and also if* q *divides both* m *and* n*, then* q *divides* p*. Two integers are relatively prime if their greatest common divisor is one. The greatest common divisor of* m *and* n *is denoted as* (m, n)*.*

There is a phenomenal and amazing theorem which relates the greatest common divisor to the smallest number in a certain set. Suppose m, n are two positive integers. Then if x, y are integers, so is $xm + yn$. Consider all integers which are of this form. Some are positive such as $1m + 1n$ and some are not. The set S in the following theorem consists of exactly those integers of this form which are positive. Then the greatest common divisor of m and n will be the smallest number in S. This is what the following theorem says.

Theorem 1.9.3 *Let* m, n *be two positive integers and define*

$$S \equiv \{xm + yn \in \mathbb{N} : x, y \in \mathbb{Z}\}.$$

Then the smallest number in S *is the greatest common divisor, denoted by* (m, n)*.*

Proof: First note that both m and n are in S so it is a nonempty set of positive integers. By well ordering, there is a smallest element of S, called $p = x_0 m + y_0 n$. Either p divides m or it does not. If p does not divide m, then by Theorem 1.8.10,

$$m = pq + r$$

where $0 < r < p$. Thus $m = (x_0 m + y_0 n)\, q + r$ and so, solving for r,

$$r = m\,(1 - x_0) + (-y_0 q)\, n \in S.$$

However, this is a contradiction because p was the smallest element of S. Thus $p|m$. Similarly $p|n$.

Now suppose q divides both m and n. Then $m = qx$ and $n = qy$ for integers, x and y. Therefore,

$$p = mx_0 + ny_0 = x_0 qx + y_0 qy = q\,(x_0 x + y_0 y)$$

showing $q|p$. Therefore, $p = (m, n)$. ■

There is a relatively simple algorithm for finding (m, n) which will be discussed now. Suppose $0 < m < n$ where m, n are integers. Also suppose the greatest common divisor is $(m, n) = d$. Then by the Euclidean algorithm, there exist integers q, r such that

$$n = qm + r, \ r < m \tag{1.1}$$

Now d divides n and m so there are numbers k, l such that $dk = m, dl = n$. From the above equation,

$$r = n - qm = dl - qdk = d(l - qk)$$

Thus d divides both m and r. If k divides both m and r, then from the equation of 1.1 it follows k also divides n. Therefore, k divides d by the definition of the greatest common divisor. Thus d is the greatest common divisor of m and r but $m + r < m + n$. This yields another pair of positive integers for which d is still the greatest common divisor but the sum of these integers is strictly smaller than the sum of the first two. Now you can do the same thing to these integers. Eventually the process must end because the sum gets strictly smaller each time it is done. It ends when there are not two positive integers produced. That is, one is a multiple of the other. At this point, the greatest common divisor is the smaller of the two numbers.

Procedure 1.9.4 *To find the greatest common divisor of m, n where $0 < m < n$, replace the pair $\{m, n\}$ with $\{m, r\}$ where $n = qm + r$ for $r < m$. This new pair of numbers has the same greatest common divisor. Do the process to this pair and continue doing this till you obtain a pair of numbers where one is a multiple of the other. Then the smaller is the sought for greatest common divisor.*

Example 1.9.5 *Find the greatest common divisor of* 165 *and* 385.

Use the Euclidean algorithm to write

$$385 = 2(165) + 55$$

Thus the next two numbers are 55 and 165. Then

$$165 = 3 \times 55$$

and so the greatest common divisor of the first two numbers is 55.

Example 1.9.6 *Find the greatest common divisor of 1237 and 4322.*

Use the Euclidean algorithm

$$4322 = 3(1237) + 611$$

Now the two new numbers are 1237,611. Then

$$1237 = 2(611) + 15$$

The two new numbers are 15,611. Then

$$611 = 40(15) + 11$$

The two new numbers are 15,11. Then

$$15 = 1(11) + 4$$

The two new numbers are 11,4

$$2(4)+3$$

The two new numbers are $4,3$. Then

$$4 = 1(3)+1$$

The two new numbers are $3,1$. Then

$$3 = 3\times 1$$

and so 1 is the greatest common divisor. Of course you could see this right away when the two new numbers were 15 and 11. Recall the process delivers numbers which have the same greatest common divisor.

This amazing theorem will now be used to prove a fundamental property of prime numbers which leads to the fundamental theorem of arithmetic, the major theorem which says every integer can be factored as a product of primes.

Theorem 1.9.7 *If p is a prime and $p|ab$ then either $p|a$ or $p|b$.*

Proof: Suppose p does not divide a. Then since p is prime, the only factors of p are 1 and p so follows $(p,a)=1$ and therefore, there exists integers, x and y such that

$$1 = ax + yp.$$

Multiplying this equation by b yields

$$b = abx + ybp.$$

Since $p|ab$, $ab = pz$ for some integer z. Therefore,

$$b = abx + ybp = pzx + ybp = p(xz+yb)$$

and this shows p divides b. ■

Theorem 1.9.8 *(Fundamental theorem of arithmetic) Let $a\in\mathbb{N}\setminus\{1\}$. Then $a=\prod_{i=1}^{n}p_i$ where p_i are all prime numbers. Furthermore, this prime factorization is unique except for the order of the factors.*

Proof: If a equals a prime number, the prime factorization clearly exists. In particular the prime factorization exists for the prime number 2. Assume this theorem is true for all $a\le n-1$. If n is a prime, then it has a prime factorization. On the other hand, if n is not a prime, then there exist two integers k and m such that $n=km$ where each of k and m are less than n. Therefore, each of these is no larger than $n-1$ and consequently, each has a prime factorization. Thus so does n. It remains to argue the prime factorization is unique except for order of the factors.

Suppose

$$\prod_{i=1}^{n}p_i = \prod_{j=1}^{m}q_j$$

where the p_i and q_j are all prime, there is no way to reorder the q_k such that $m=n$ and $p_i=q_i$ for all i, and $n+m$ is the smallest positive integer such that this happens. Then by Theorem 1.9.7, $p_1|q_j$ for some j. Since these are prime numbers this requires $p_1=q_j$. Reordering if necessary it can be assumed that $q_j=q_1$. Then dividing both sides by $p_1=q_1$,

$$\prod_{i=1}^{n-1}p_{i+1} = \prod_{j=1}^{m-1}q_{j+1}.$$

Since $n+m$ was as small as possible for the theorem to fail, it follows that $n-1=m-1$ and the prime numbers, $q_2,\cdots,q_m$ can be reordered in such a way that $p_k=q_k$ for all $k=2,\cdots,n$. Hence $p_i=q_i$ for all i because it was already argued that $p_1=q_1$, and this results in a contradiction. ■

1.10 Systems Of Equations

Sometimes it is necessary to solve systems of equations. For example the problem could be to find x and y such that

$$x+y=7 \text{ and } 2x-y=8. \tag{1.2}$$

The set of ordered pairs, (x,y) which solve both equations is called the solution set. For example, you can see that $(5,2)=(x,y)$ is a solution to the above system. To solve this, note that the solution set does not change if any equation is replaced by a non zero multiple of itself. It also does not change if one equation is replaced by itself added to a multiple of the other equation. For example, x and y solve the above system if and only if x and y solve the system

$$x+y=7, \overbrace{2x-y+(-2)(x+y)=8+(-2)(7)}^{-3y=-6}. \tag{1.3}$$

The second equation was replaced by -2 times the first equation added to the second. Thus the solution is $y=2$, from $-3y=-6$ and now, knowing $y=2$, it follows from the other equation that $x+2=7$ and so $x=5$.

Why exactly does the replacement of one equation with a multiple of another added to it not change the solution set? The two equations of 1.2 are of the form

$$E_1=f_1, E_2=f_2 \tag{1.4}$$

where E_1 and E_2 are expressions involving the variables. The claim is that if a is a number, then 1.4 has the same solution set as

$$E_1=f_1,\ E_2+aE_1=f_2+af_1. \tag{1.5}$$

Why is this?

If (x, y) solves 1.4 then it solves the first equation in 1.5. Also, it satisfies $aE_1 = af_1$ and so, since it also solves $E_2 = f_2$ it must solve the second equation in 1.5. If (x, y) solves 1.5 then it solves the first equation of 1.4. Also $aE_1 = af_1$ and it is given that the second equation of 1.5 is verified. Therefore, $E_2 = f_2$ and it follows (x, y) is a solution of the second equation in 1.4. This shows the solutions to 1.4 and 1.5 are exactly the same which means they have the same solution set. Of course the same reasoning applies with no change if there are many more variables than two and many more equations than two. It is still the case that when one equation is replaced with a multiple of another one added to itself, the solution set of the whole system does not change.

The other thing which does not change the solution set of a system of equations consists of listing the equations in a different order. Here is another example.

Example 1.10.1 *Find the solutions to the system,*

$$\begin{aligned} x+3y+6z &= 25 \\ 2x+7y+14z &= 58 \\ 2y+5z &= 19 \end{aligned} \tag{1.6}$$

To solve this system replace the second equation by (-2) times the first equation added to the second. This yields. the system

$$\begin{aligned} x+3y+6z &= 25 \\ y+2z &= 8 \\ 2y+5z &= 19 \end{aligned} \tag{1.7}$$

Now take (-2) times the second and add to the third. More precisely, replace the third equation with (-2) times the second added to the third. This yields the system

$$\begin{aligned} x+3y+6z &= 25 \\ y+2z &= 8 \\ z &= 3 \end{aligned} \tag{1.8}$$

At this point, you can tell what the solution is. This system has the same solution as the original system and in the above, $z = 3$. Then using this in the second equation, it follows $y + 6 = 8$ and so $y = 2$. Now using this in the top equation yields $x + 6 + 18 = 25$ and so $x = 1$.

This process is not really much different from what you have always done in solving a single equation. For example, suppose you wanted to solve $2x + 5 = 3x - 6$. You did the same thing to both sides of the equation thus preserving the solution set until you obtained an equation which was simple enough to give the answer. In this case, you would add $-2x$ to both sides and then add 6 to both sides. This yields $x = 11$.

In 1.8 you could have continued as follows. Add (-2) times the bottom equation to the middle and then add (-6) times the bottom to the top. This yields

$$\begin{aligned} x+3y &= 19 \\ y &= 6 \\ z &= 3 \end{aligned}$$

Now add (-3) times the second to the top. This yields

$$\begin{aligned} x &= 1 \\ y &= 6 \\ z &= 3 \end{aligned}\ ,$$

a system which has the same solution set as the original system.

It is foolish to write the variables every time you do these operations. It is easier to write the system 1.6 as the following "augmented matrix"

$$\begin{pmatrix} 1 & 3 & 6 & 25 \\ 2 & 7 & 14 & 58 \\ 0 & 2 & 5 & 19 \end{pmatrix}.$$

It has exactly the same information as the original system but here it is understood there is an x column, $\begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}$, a y column, $\begin{pmatrix} 3 \\ 7 \\ 2 \end{pmatrix}$ and a z column, $\begin{pmatrix} 6 \\ 14 \\ 5 \end{pmatrix}$. The rows correspond to the equations in the system. Thus the top row in the augmented matrix corresponds to the equation,

$$x + 3y + 6z = 25.$$

Now when you replace an equation with a multiple of another equation added to itself, you are just taking a row of this augmented matrix and replacing it with a multiple of another row added to it. Thus the first step in solving 1.6 would be to take (-2) times the first row of the augmented matrix above and add it to the second row,

$$\begin{pmatrix} 1 & 3 & 6 & 25 \\ 0 & 1 & 2 & 8 \\ 0 & 2 & 5 & 19 \end{pmatrix}.$$

Note how this corresponds to 1.7. Next take (-2) times the second row and add to the third,

$$\begin{pmatrix} 1 & 3 & 6 & 25 \\ 0 & 1 & 2 & 8 \\ 0 & 0 & 1 & 3 \end{pmatrix}$$

which is the same as 1.8. You get the idea I hope. Write the system as an augmented matrix and follow the procedure of either switching rows, multiplying a row by a non zero number, or replacing a row by a multiple of another row added to it. Each of these operations leaves the solution set unchanged. These operations are called row operations.

Definition 1.10.2 *The row operations consist of the following*

1. *Switch two rows.*
2. *Multiply a row by a nonzero number.*
3. *Replace a row by a multiple of another row added to it.*

It is important to observe that any row operation can be "undone" by another inverse row operation. For example, if $\mathbf{r}_1, \mathbf{r}_2$ are two rows, and $\mathbf{r}_2$ is replaced with $\mathbf{r}_2' = \alpha \mathbf{r}_1 + \mathbf{r}_2$ using row operation 3, then you could get back to where you started by replacing the row $\mathbf{r}_2'$ with $-\alpha$ times $\mathbf{r}_1$ and adding to $\mathbf{r}_2'$. In the case of operation 2, you would simply multiply the row that was changed by the inverse of the scalar which multiplied it in the first place, and in the case of row operation 1, you would just make the same switch again and you would be back to where you started. In each case, the row operation which undoes what was done is called the **inverse row operation.**

Example 1.10.3 *Give the complete solution to the system of equations,* $5x+10y-7z = -2$, $2x + 4y - 3z = -1$, *and* $3x + 6y + 5z = 9$.

The augmented matrix for this system is

$$\begin{pmatrix} 2 & 4 & -3 & -1 \\ 5 & 10 & -7 & -2 \\ 3 & 6 & 5 & 9 \end{pmatrix}$$

Multiply the second row by 2, the first row by 5, and then take (-1) times the first row and add to the second. Then multiply the first row by $1/5$. This yields

$$\begin{pmatrix} 2 & 4 & -3 & -1 \\ 0 & 0 & 1 & 1 \\ 3 & 6 & 5 & 9 \end{pmatrix}$$

Now, combining some row operations, take (-3) times the first row and add this to 2 times the last row and replace the last row with this. This yields.

$$\begin{pmatrix} 2 & 4 & -3 & -1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 21 \end{pmatrix}.$$

Putting in the variables, the last two rows say $z = 1$ and $z = 21$. This is impossible so the last system of equations determined by the above augmented matrix has no solution. However, it has the same solution set as the first system of equations. This shows there is no solution to the three given equations. When this happens, the system is called inconsistent.

This should not be surprising that something like this can take place. It can even happen for one equation in one variable. Consider for example, $x = x+1$. There is clearly no solution to this.

Example 1.10.4 *Give the complete solution to the system of equations,* $3x - y - 5z = 9$, $y - 10z = 0$, *and* $-2x + y = -6$.

The augmented matrix of this system is

$$\begin{pmatrix} 3 & -1 & -5 & 9 \\ 0 & 1 & -10 & 0 \\ -2 & 1 & 0 & -6 \end{pmatrix}$$

Replace the last row with 2 times the top row added to 3 times the bottom row. This gives

$$\begin{pmatrix} 3 & -1 & -5 & 9 \\ 0 & 1 & -10 & 0 \\ 0 & 1 & -10 & 0 \end{pmatrix}$$

Next take -1 times the middle row and add to the bottom.

$$\begin{pmatrix} 3 & -1 & -5 & 9 \\ 0 & 1 & -10 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Take the middle row and add to the top and then divide the top row which results by 3.

$$\begin{pmatrix} 1 & 0 & -5 & 3 \\ 0 & 1 & -10 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

This says $y = 10z$ and $x = 3 + 5z$. Apparently z can equal any number. Therefore, the solution set of this system is $x = 3 + 5t, y = 10t$, and $z = t$ where t is completely arbitrary. The system has an infinite set of solutions and this is a good description of the solutions.

This is what it is all about, finding the solutions to the system.

Definition 1.10.5 *Since $z = t$ where t is arbitrary, the variable z is called a* ***free variable.***

The phenomenon of an infinite solution set occurs in equations having only one variable also. For example, consider the equation $x = x$. It doesn't matter what x equals.

Definition 1.10.6 *A system of linear equations is a list of equations,*

$$\sum_{j=1}^{n} a_{ij} x_j = f_j, \;\; i = 1, 2, 3, \cdots, m$$

where a_{ij} are numbers, f_j is a number, and it is desired to find $(x_1, \cdots, x_n)$ solving each of the equations listed.

As illustrated above, such a system of linear equations may have a unique solution, no solution, or infinitely many solutions. It turns out these are the only three cases which can occur for linear systems. Furthermore, you do exactly the same things to solve any linear system. You write the augmented matrix and do row operations until you get a simpler system in which it is possible to see the solution. All is based on the observation that the row operations do not change the solution set. You can have more equations than variables, fewer equations than variables, etc. It doesn't matter. You always set up the augmented matrix and go to work on it. These things are all the same.

Example 1.10.7 *Give the complete solution to the system of equations,* $-41x+15y=168$, $109x-40y=-447$, $-3x+y=12$, *and* $2x+z=-1$.

The augmented matrix is

$$\begin{pmatrix} -41 & 15 & 0 & 168 \\ 109 & -40 & 0 & -447 \\ -3 & 1 & 0 & 12 \\ 2 & 0 & 1 & -1 \end{pmatrix}.$$

To solve this multiply the top row by 109, the second row by 41, add the top row to the second row, and multiply the top row by 1/109. Note how this process combined several row operations. This yields

$$\begin{pmatrix} -41 & 15 & 0 & 168 \\ 0 & -5 & 0 & -15 \\ -3 & 1 & 0 & 12 \\ 2 & 0 & 1 & -1 \end{pmatrix}.$$

Next take 2 times the third row and replace the fourth row by this added to 3 times the fourth row. Then take (-41) times the third row and replace the first row by this added to 3 times the first row. Then switch the third and the first rows. This yields

$$\begin{pmatrix} 123 & -41 & 0 & -492 \\ 0 & -5 & 0 & -15 \\ 0 & 4 & 0 & 12 \\ 0 & 2 & 3 & 21 \end{pmatrix}.$$

Take $-1/2$ times the third row and add to the bottom row. Then take 5 times the third row and add to four times the second. Finally take 41 times the third row and add to 4 times the top row. This yields

$$\begin{pmatrix} 492 & 0 & 0 & -1476 \\ 0 & 0 & 0 & 0 \\ 0 & 4 & 0 & 12 \\ 0 & 0 & 3 & 15 \end{pmatrix}$$

It follows $x=\frac{-1476}{492}=-3, y=3$ and $z=5$.

You should practice solving systems of equations. Here are some exercises.

1.11 Exercises

1. Give the complete solution to the system of equations, $3x-y+4z=6$, $y+8z=0$, and $-2x+y=-4$.

2. Give the complete solution to the system of equations, $x+3y+3z=3$, $3x+2y+z=9$, and $-4x+z=-9$.

3. Consider the system $-5x + 2y - z = 0$ and $-5x - 2y - z = 0$. Both equations equal zero and so $-5x + 2y - z = -5x - 2y - z$ which is equivalent to $y = 0$. Thus x and z can equal anything. But when $x = 1$, $z = -4$, and $y = 0$ are plugged in to the equations, it doesn't work. Why?

4. Give the complete solution to the system of equations, $x+2y+6z = 5, 3x+2y+6z = 7$ $,-4x + 5y + 15z = -7$.

5. Give the complete solution to the system of equations

$$\begin{aligned} x + 2y + 3z &= 5, 3x + 2y + z = 7, \\ -4x + 5y + z &= -7, x + 3z = 5. \end{aligned}$$

6. Give the complete solution of the system of equations,

$$\begin{aligned} x + 2y + 3z &= 5,\ 3x + 2y + 2z = 7 \\ -4x + 5y + 5z &= -7,\ x = 5 \end{aligned}$$

7. Give the complete solution of the system of equations

$$\begin{aligned} x + y + 3z &= 2,\ 3x - y + 5z = 6 \\ -4x + 9y + z &= -8,\ x + 5y + 7z = 2 \end{aligned}$$

8. Determine a such that there are infinitely many solutions and then find them. Next determine a such that there are no solutions. Finally determine which values of a correspond to a unique solution. The system of equations for the unknown variables x, y, z is

$$\begin{gathered} 3za^2 - 3a + x + y + 1 = 0 \\ 3x - a - y + z\left(a^2 + 4\right) - 5 = 0 \\ za^2 - a - 4x + 9y + 9 = 0 \end{gathered}$$

9. Find the solutions to the following system of equations for x, y, z, w.

$$y + z = 2, z + w = 0, y - 4z - 5w = 2, 2y + z - w = 4$$

10. Find all solutions to the following equations.

$$\begin{aligned} x + y + z &= 2,\ z + w = 0, \\ 2x + 2y + z - w &= 4,\ x + y - 4z - 5z = 2 \end{aligned}$$

1.12 $\mathbb{F}^n$

The notation, $\mathbb{C}^n$ refers to the collection of ordered lists of n complex numbers. Since every real number is also a complex number, this simply generalizes the usual notion of $\mathbb{R}^n$, the collection of all ordered lists of n real numbers. In order to avoid worrying about whether it is real or complex numbers which are being referred to, the symbol $\mathbb{F}$ will be used. If it is not clear, always pick $\mathbb{C}$. More generally, $\mathbb{F}^n$ refers to the ordered lists of n elements of $\mathbb{F}^n$.

Definition 1.12.1 *Define* $\mathbb{F}^n \equiv \{(x_1, \cdots, x_n) : x_j \in \mathbb{F} \text{ for } j = 1, \cdots, n\}$. $(x_1, \cdots, x_n) = (y_1, \cdots, y_n)$ *if and only if for all* $j = 1, \cdots, n$, $x_j = y_j$. *When* $(x_1, \cdots, x_n) \in \mathbb{F}^n$, *it is* *conventional to denote* $(x_1, \cdots, x_n)$ *by the single bold face letter* $\mathbf{x}$. *The numbers* x_j *are called the coordinates. The set*

$$\{(0, \cdots, 0, t, 0, \cdots, 0) : t \in \mathbb{F}\}$$

for t *in the* i^{th} *slot is called the* i^{th} *coordinate axis. The point* $\mathbf{0} \equiv (0, \cdots, 0)$ *is called the origin.*

Thus $(1, 2, 4i) \in \mathbb{F}^3$ and $(2, 1, 4i) \in \mathbb{F}^3$ but $(1, 2, 4i) \neq (2, 1, 4i)$ because, even though the same numbers are involved, they don't match up. In particular, the first entries are not equal.

1.13 Algebra in $\mathbb{F}^n$

There are two algebraic operations done with elements of $\mathbb{F}^n$. One is addition and the other is multiplication by numbers, called scalars. In the case of $\mathbb{C}^n$ the scalars are complex numbers while in the case of $\mathbb{R}^n$ the only allowed scalars are real numbers. Thus, the scalars always come from $\mathbb{F}$ in either case.

Definition 1.13.1 *If* $\mathbf{x} \in \mathbb{F}^n$ *and* $a \in \mathbb{F}$, *also called a scalar, then* $a\mathbf{x} \in \mathbb{F}^n$ *is defined by*

$$a\mathbf{x} = a(x_1, \cdots, x_n) \equiv (ax_1, \cdots, ax_n). \tag{1.9}$$

This is known as scalar multiplication. If $\mathbf{x}, \mathbf{y} \in \mathbb{F}^n$ *then* $\mathbf{x} + \mathbf{y} \in \mathbb{F}^n$ *and is defined by*

$$\begin{aligned} \mathbf{x} + \mathbf{y} &= (x_1, \cdots, x_n) + (y_1, \cdots, y_n) \\ &\equiv (x_1 + y_1, \cdots, x_n + y_n) \end{aligned} \tag{1.10}$$

With this definition, the algebraic properties satisfy the conclusions of the following theorem.

Theorem 1.13.2 *For* $\mathbf{v}, \mathbf{w} \in \mathbb{F}^n$ *and* α, β *scalars, (real numbers), the following hold.*

$$\mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v}, \tag{1.11}$$

the commutative law of addition,

$$(\mathbf{v} + \mathbf{w}) + \mathbf{z} = \mathbf{v} + (\mathbf{w} + \mathbf{z}), \tag{1.12}$$

the associative law for addition,

$$\mathbf{v} + \mathbf{0} = \mathbf{v}, \tag{1.13}$$

the existence of an additive identity,

$$\mathbf{v} + (-\mathbf{v}) = \mathbf{0}, \tag{1.14}$$

the existence of an additive inverse, Also

$$\alpha(\mathbf{v} + \mathbf{w}) = \alpha\mathbf{v} + \alpha\mathbf{w}, \tag{1.15}$$

$$(\alpha + \beta)\mathbf{v} = \alpha\mathbf{v} + \beta\mathbf{v}, \tag{1.16}$$

$$\alpha(\beta\mathbf{v}) = \alpha\beta(\mathbf{v}), \tag{1.17}$$

$$1\mathbf{v} = \mathbf{v}. \tag{1.18}$$

In the above $\mathbf{0} = (0, \cdots, 0)$.

You should verify that these properties all hold. As usual subtraction is defined as $\mathbf{x} - \mathbf{y} \equiv \mathbf{x} + (-\mathbf{y})$. The conclusions of the above theorem are called the vector space axioms.

1.14 Exercises

1. Verify all the properties 1.11-1.18.
2. Compute $5(1, 2 + 3i, 3, -2) + 6(2 - i, 1, -2, 7)$.
3. Draw a picture of the points in $\mathbb{R}^2$ which are determined by the following ordered pairs.
   (a) $(1, 2)$
   (b) $(-2, -2)$
   (c) $(-2, 3)$

(d) $(2, -5)$

4. Does it make sense to write $(1, 2) + (2, 3, 1)$? Explain.

5. Draw a picture of the points in $\mathbb{R}^3$ which are determined by the following ordered triples. If you have trouble drawing this, describe it in words.

   (a) $(1, 2, 0)$
   (b) $(-2, -2, 1)$
   (c) $(-2, 3, -2)$

1.15 The Inner Product In $\mathbb{F}^n$

When $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$, there is something called an inner product. In case of $\mathbb{R}$ it is also called the dot product. This is also often referred to as the scalar product.

Definition 1.15.1 *Let* $\mathbf{a}, \mathbf{b} \in \mathbb{F}^n$ *define* $\mathbf{a} \cdot \mathbf{b}$ *as*

$$\mathbf{a} \cdot \mathbf{b} \equiv \sum_{k=1}^{n} a_k \overline{b}_k.$$

With this definition, there are several important properties satisfied by the inner product. In the statement of these properties, α and β will denote scalars and $\mathbf{a}, \mathbf{b}, \mathbf{c}$ will denote vectors or in other words, points in $\mathbb{F}^n$.

Proposition 1.15.2 *The inner product satisfies the following properties.*

$$\mathbf{a} \cdot \mathbf{b} = \overline{\mathbf{b} \cdot \mathbf{a}} \tag{1.19}$$

$$\mathbf{a} \cdot \mathbf{a} \geq 0 \text{ and equals zero if and only if } \mathbf{a} = \mathbf{0} \tag{1.20}$$

$$(\alpha \mathbf{a} + \beta \mathbf{b}) \cdot \mathbf{c} = \alpha (\mathbf{a} \cdot \mathbf{c}) + \beta (\mathbf{b} \cdot \mathbf{c}) \tag{1.21}$$

$$\mathbf{c} \cdot (\alpha \mathbf{a} + \beta \mathbf{b}) = \overline{\alpha} (\mathbf{c} \cdot \mathbf{a}) + \overline{\beta} (\mathbf{c} \cdot \mathbf{b}) \tag{1.22}$$

$$|\mathbf{a}|^2 = \mathbf{a} \cdot \mathbf{a} \tag{1.23}$$

You should verify these properties. Also be sure you understand that 1.22 follows from the first three and is therefore redundant. It is listed here for the sake of convenience.

Example 1.15.3 *Find* $(1, 2, 0, -1) \cdot (0, i, 2, 3)$.

This equals $0 + 2(-i) + 0 + -3 = -3 - 2i$

The Cauchy Schwarz inequality takes the following form in terms of the inner product. I will prove it using only the above axioms for the inner product.

Theorem 1.15.4 *The inner product satisfies the inequality*

$$|\mathbf{a} \cdot \mathbf{b}| \leq |\mathbf{a}| \, |\mathbf{b}| \, . \tag{1.24}$$

Furthermore equality is obtained if and only if one of $\mathbf{a}$ *or* $\mathbf{b}$ *is a scalar multiple of the other.*

Proof: First define $\theta \in \mathbb{C}$ such that

$$\overline{\theta} (\mathbf{a} \cdot \mathbf{b}) = |\mathbf{a} \cdot \mathbf{b}|, |\theta| = 1,$$

and define a function of $t \in \mathbb{R}$

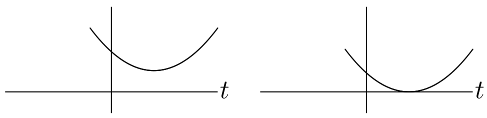
$$f(t) = (\mathbf{a} + t\theta\mathbf{b}) \cdot (\mathbf{a} + t\theta\mathbf{b}).$$

Then by 1.20, $f(t) \geq 0$ for all $t \in \mathbb{R}$. Also from 1.21,1.22,1.19, and 1.23

$$\begin{aligned} f(t) &= \mathbf{a} \cdot (\mathbf{a} + t\theta\mathbf{b}) + t\theta\mathbf{b} \cdot (\mathbf{a} + t\theta\mathbf{b}) \\ &= \mathbf{a} \cdot \mathbf{a} + t\overline{\theta} (\mathbf{a} \cdot \mathbf{b}) + t\theta (\mathbf{b} \cdot \mathbf{a}) + t^2 |\theta|^2 \mathbf{b} \cdot \mathbf{b} \end{aligned}$$

$$= |\mathbf{a}|^2 + 2t \operatorname{Re} \overline{\theta} (\mathbf{a} \cdot \mathbf{b}) + |\mathbf{b}|^2 t^2 = |\mathbf{a}|^2 + 2t |\mathbf{a} \cdot \mathbf{b}| + |\mathbf{b}|^2 t^2$$

Now if $|\mathbf{b}|^2 = 0$ it must be the case that $\mathbf{a} \cdot \mathbf{b} = 0$ because otherwise, you could pick large negative values of t and violate $f(t) \geq 0$. Therefore, in this case, the Cauchy Schwarz inequality holds. In the case that $|\mathbf{b}| \neq 0$, $y = f(t)$ is a polynomial which opens up and therefore, if it is always nonnegative, its graph is like that illustrated in the following picture



Then the quadratic formula requires that

$$\overbrace{4 |\mathbf{a} \cdot \mathbf{b}|^2 - 4 |\mathbf{a}|^2 |\mathbf{b}|^2}^{\text{The discriminant}} \leq 0$$

since otherwise the function, $f(t)$ would have two real zeros and would necessarily have a graph which dips below the t axis. This proves 1.24.

It is clear from the axioms of the inner product that equality holds in 1.24 whenever one of the vectors is a scalar multiple of the other. It only remains to verify this is the only way equality can occur. If either vector equals zero, then equality is obtained in 1.24 so it can be assumed both vectors are non zero. Then if equality is achieved, it follows $f(t)$ has exactly one real zero because the discriminant vanishes. Therefore, for some value of $t, \mathbf{a} + t\theta\mathbf{b} = \mathbf{0}$ showing that $\mathbf{a}$ is a multiple of $\mathbf{b}$. ■

You should note that the entire argument was based only on the properties of the inner product listed in 1.19 - 1.23. This means that whenever something satisfies these properties, the Cauchy Schwartz inequality holds. There are many other instances of these properties besides vectors in $\mathbb{F}^n$. Also note that 1.24 holds if 1.20 is simplified to $\mathbf{a} \cdot \mathbf{a} \geq 0$.

The Cauchy Schwartz inequality allows a proof of the triangle inequality for distances in $\mathbb{F}^n$ in much the same way as the triangle inequality for the absolute value.

Theorem 1.15.5 *(Triangle inequality) For* $\mathbf{a}, \mathbf{b} \in \mathbb{F}^n$

$$|\mathbf{a} + \mathbf{b}| \leq |\mathbf{a}| + |\mathbf{b}| \tag{1.25}$$

and equality holds if and only if one of the vectors is a nonnegative scalar multiple of the other. Also

$$||\mathbf{a}| - |\mathbf{b}|| \leq |\mathbf{a} - \mathbf{b}| \tag{1.26}$$

Proof: By properties of the inner product and the Cauchy Schwartz inequality,

$$|\mathbf{a}+\mathbf{b}|^2 = (\mathbf{a}+\mathbf{b})\cdot(\mathbf{a}+\mathbf{b}) = (\mathbf{a}\cdot\mathbf{a}) + (\mathbf{a}\cdot\mathbf{b}) + (\mathbf{b}\cdot\mathbf{a}) + (\mathbf{b}\cdot\mathbf{b})$$

$$= |\mathbf{a}|^2 + 2\,\mathrm{Re}\,(\mathbf{a}\cdot\mathbf{b}) + |\mathbf{b}|^2 \leq |\mathbf{a}|^2 + 2\,|\mathbf{a}\cdot\mathbf{b}| + |\mathbf{b}|^2$$

$$\leq |\mathbf{a}|^2 + 2\,|\mathbf{a}|\,|\mathbf{b}| + |\mathbf{b}|^2 = (|\mathbf{a}| + |\mathbf{b}|)^2 .$$

Taking square roots of both sides you obtain 1.25.

It remains to consider when equality occurs. If either vector equals zero, then that vector equals zero times the other vector and the claim about when equality occurs is verified. Therefore, it can be assumed both vectors are nonzero. To get equality in the second inequality above, Theorem 1.15.4 implies one of the vectors must be a multiple of the other. Say $\mathbf{b} = \alpha\mathbf{a}$. Also, to get equality in the first inequality, $(\mathbf{a}\cdot\mathbf{b})$ must be a nonnegative real number. Thus

$$0 \leq (\mathbf{a}\cdot\mathbf{b}) = (\mathbf{a}\cdot\alpha\mathbf{a}) = \overline{\alpha}\,|\mathbf{a}|^2 .$$

Therefore, α must be a real number which is nonnegative.

To get the other form of the triangle inequality,

$$\mathbf{a} = \mathbf{a} - \mathbf{b} + \mathbf{b}$$

so

$$|\mathbf{a}| = |\mathbf{a} - \mathbf{b} + \mathbf{b}| \leq |\mathbf{a} - \mathbf{b}| + |\mathbf{b}| .$$

Therefore,

$$|\mathbf{a}| - |\mathbf{b}| \leq |\mathbf{a} - \mathbf{b}| \tag{1.27}$$

Similarly,

$$|\mathbf{b}| - |\mathbf{a}| \leq |\mathbf{b} - \mathbf{a}| = |\mathbf{a} - \mathbf{b}| . \tag{1.28}$$

It follows from 1.27 and 1.28 that 1.26 holds. This is because $||\mathbf{a}| - |\mathbf{b}||$ equals the left side of either 1.27 or 1.28 and either way, $||\mathbf{a}| - |\mathbf{b}|| \leq |\mathbf{a} - \mathbf{b}|$. ■

1.16 What Is Linear Algebra?

The above preliminary considerations form the necessary scaffolding upon which linear algebra is built. Linear algebra is the study of a certain algebraic structure called a vector space described in a special case in Theorem 1.13.2 and in more generality below along with special functions known as linear transformations. These linear transformations preserve certain algebraic properties.

A good argument could be made that linear algebra is the most useful subject in all of mathematics and that it exceeds even courses like calculus in its significance. It is used extensively in applied mathematics and engineering. Continuum mechanics, for example, makes use of topics from linear algebra in defining things like the strain and in determining appropriate constitutive laws. It is fundamental in the study of statistics. For example, principal component analysis is really based on the singular value decomposition discussed in this book. It is also fundamental in pure mathematics areas like number theory, functional analysis, geometric measure theory, and differential geometry. Even calculus cannot be correctly understood without it. For example, the derivative of a function of many variables is an example of a linear transformation, and this is the way it must be understood as soon as you consider functions of more than one variable.

1.17 Exercises

1. Show that $(\mathbf{a} \cdot \mathbf{b}) = \frac{1}{4}\left[|\mathbf{a}+\mathbf{b}|^2 - |\mathbf{a}-\mathbf{b}|^2\right]$.

2. Prove from the axioms of the inner product the parallelogram identity, $|\mathbf{a}+\mathbf{b}|^2 + |\mathbf{a}-\mathbf{b}|^2 = 2|\mathbf{a}|^2 + 2|\mathbf{b}|^2$.

3. For $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$, define $\mathbf{a} \cdot \mathbf{b} \equiv \sum_{k=1}^n \beta_k a_k b_k$ where $\beta_k > 0$ for each k. Show this satisfies the axioms of the inner product. What does the Cauchy Schwarz inequality say in this case.

4. In Problem 3 above, suppose you only know $\beta_k \geq 0$. Does the Cauchy Schwarz inequality still hold? If so, prove it.

5. Let f, g be continuous functions and define

$$f \cdot g \equiv \int_0^1 f(t)\,\overline{g(t)}dt$$

show this satisfies the axioms of a inner product if you think of continuous functions in the place of a vector in $\mathbb{F}^n$. What does the Cauchy Schwarz inequality say in this case?

6. Show that if f is a real valued continuous function,

$$\left(\int_a^b f(t)\, dt \right)^2 \leq (b-a) \int_a^b f(t)^2\, dt.$$

Linear Transformations

2.1 Matrices

You have now solved systems of equations by writing them in terms of an augmented matrix and then doing row operations on this augmented matrix. It turns out that such rectangular arrays of numbers are important from many other different points of view. Numbers are also called scalars. In general, scalars are just elements of some field. However, in the first part of this book, the field will typically be either the real numbers or the complex numbers.

A matrix is a rectangular array of numbers. Several of them are referred to as matrices. For example, here is a matrix.

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 2 & 8 & 7 \\ 6 & -9 & 1 & 2 \end{pmatrix}$$

This matrix is a 3×4 matrix because there are three rows and four columns. The first row is $(1\ 2\ 3\ 4)$, the second row is $(5\ 2\ 8\ 7)$ and so forth. The first column is $\begin{pmatrix} 1 \\ 5 \\ 6 \end{pmatrix}$. The convention in dealing with matrices is to always list the rows first and then the columns. Also, you can remember the columns are like columns in a Greek temple. They stand up right while the rows just lay there like rows made by a tractor in a plowed field. Elements of the matrix are identified according to position in the matrix. For example, 8 is in position $2, 3$ because it is in the second row and the third column. You might remember that you always list the rows before the columns by using the phrase **Row**man **C**atholic. The symbol, (a_{ij}) refers to a matrix in which the i denotes the row and the j denotes the column. Using this notation on the above matrix, $a_{23} = 8, a_{32} = -9, a_{12} = 2$, etc.

There are various operations which are done on matrices. They can sometimes be added, multiplied by a scalar and sometimes multiplied. To illustrate scalar multiplication, consider the following example.

$$3 \begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 2 & 8 & 7 \\ 6 & -9 & 1 & 2 \end{pmatrix} = \begin{pmatrix} 3 & 6 & 9 & 12 \\ 15 & 6 & 24 & 21 \\ 18 & -27 & 3 & 6 \end{pmatrix}.$$

The new matrix is obtained by multiplying every entry of the original matrix by the given scalar. If A is an $m \times n$ matrix $-A$ is defined to equal $(-1)\, A$.

Two matrices which are the same size can be added. When this is done, the result is the matrix which is obtained by adding corresponding entries. Thus

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 2 \end{pmatrix} + \begin{pmatrix} -1 & 4 \\ 2 & 8 \\ 6 & -4 \end{pmatrix} = \begin{pmatrix} 0 & 6 \\ 5 & 12 \\ 11 & -2 \end{pmatrix}.$$

Two matrices are equal exactly when they are the same size and the corresponding entries are identical. Thus

$$\begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \neq \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

because they are different sizes. As noted above, you write (c_{ij}) for the matrix C whose ij^{th} entry is c_{ij}. In doing arithmetic with matrices you must define what happens in terms of the c_{ij} sometimes called the entries of the matrix or the components of the matrix.

The above discussion stated for general matrices is given in the following definition.

Definition 2.1.1 *Let $A = (a_{ij})$ and $B = (b_{ij})$ be two $m \times n$ matrices. Then $A + B = C$ where*

$$C = (c_{ij})$$

for $c_{ij} = a_{ij} + b_{ij}$. Also if x is a scalar,

$$xA = (c_{ij})$$

where $c_{ij} = xa_{ij}$. The number A_{ij} will typically refer to the ij^{th} entry of the matrix A. The zero matrix, denoted by 0 will be the matrix consisting of all zeros.

Do not be upset by the use of the subscripts, ij. The expression $c_{ij} = a_{ij} + b_{ij}$ is just saying that you add corresponding entries to get the result of summing two matrices as discussed above.

Note that there are 2×3 zero matrices, 3×4 zero matrices, etc. In fact for every size there is a zero matrix.

With this definition, the following properties are all obvious but you should verify all of these properties are valid for A, B, and C, $m \times n$ matrices and 0 an $m \times n$ zero matrix,

$$A + B = B + A, \tag{2.1}$$

the commutative law of addition,

$$(A + B) + C = A + (B + C), \tag{2.2}$$

the associative law for addition,

$$A + 0 = A, \tag{2.3}$$

the existence of an additive identity,

$$A + (-A) = 0, \tag{2.4}$$

the existence of an additive inverse. Also, for α, β scalars, the following also hold.

$$\alpha (A + B) = \alpha A + \alpha B, \tag{2.5}$$

$$(\alpha + \beta) A = \alpha A + \beta A, \tag{2.6}$$

$$\alpha (\beta A) = \alpha\beta (A), \tag{2.7}$$

$$1A = A. \tag{2.8}$$

The above properties, 2.1 - 2.8 are known as the vector space axioms and the fact that the $m \times n$ matrices satisfy these axioms is what is meant by saying this set of matrices with addition and scalar multiplication as defined above forms a vector space.

Definition 2.1.2 *Matrices which are $n \times 1$ or $1 \times n$ are especially called vectors and are often denoted by a bold letter. Thus*

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

is an $n \times 1$ matrix also called a column vector while a $1 \times n$ matrix of the form $(x_1 \cdots x_n)$ is referred to as a row vector.

All the above is fine, but the real reason for considering matrices is that they can be multiplied. This is where things quit being banal.

First consider the problem of multiplying an $m \times n$ matrix by an $n \times 1$ column vector. Consider the following example

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \begin{pmatrix} 7 \\ 8 \\ 9 \end{pmatrix} = ?$$

It equals

$$7 \begin{pmatrix} 1 \\ 4 \end{pmatrix} + 8 \begin{pmatrix} 2 \\ 5 \end{pmatrix} + 9 \begin{pmatrix} 3 \\ 6 \end{pmatrix}$$

Thus it is what is called a **linear combination** of the columns. These will be discussed more later. Motivated by this example, here is the definition of how to multiply an $m \times n$ matrix by an $n \times 1$ matrix. (vector)

Definition 2.1.3 *Let $A = A_{ij}$ be an $m \times n$ matrix and let $\mathbf{v}$ be an $n \times 1$ matrix,*

$$\mathbf{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}, \ A = (\mathbf{a}_1, \cdots, \mathbf{a}_n)$$

where $\mathbf{a}_i$ is an $m \times 1$ vector. Then $A\mathbf{v}$, written as

$$\begin{pmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix},$$

is the $m \times 1$ column vector which equals the following linear combination of the columns.

$$v_1 \mathbf{a}_1 + v_2 \mathbf{a}_2 + \cdots + v_n \mathbf{a}_n \equiv \sum_{j=1}^{n} v_j \mathbf{a}_j \tag{2.9}$$

If the j^{th} column of A is

$$\begin{pmatrix} A_{1j} \\ A_{2j} \\ \vdots \\ A_{mj} \end{pmatrix}$$

then 2.9 takes the form

$$v_1 \begin{pmatrix} A_{11} \\ A_{21} \\ \vdots \\ A_{m1} \end{pmatrix} + v_2 \begin{pmatrix} A_{12} \\ A_{22} \\ \vdots \\ A_{m2} \end{pmatrix} + \cdots + v_n \begin{pmatrix} A_{1n} \\ A_{2n} \\ \vdots \\ A_{mn} \end{pmatrix}$$

Thus the i^{th} *entry of* $A\mathbf{v}$ *is* $\sum_{j=1}^{n} A_{ij}v_j$. *Note that multiplication by an* $m \times n$ *matrix takes an* $n \times 1$ *matrix, and produces an* $m \times 1$ *matrix (vector).*

Here is another example.

Example 2.1.4 *Compute*

$$\begin{pmatrix} 1 & 2 & 1 & 3 \\ 0 & 2 & 1 & -2 \\ 2 & 1 & 4 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 0 \\ 1 \end{pmatrix}.$$

First of all, this is of the form $(3 \times 4)(4 \times 1)$ and so the result should be a (3×1). Note how the inside numbers cancel. To get the entry in the second row and first and only column, compute

$$\begin{aligned}\sum_{k=1}^{4} a_{2k}v_k &= a_{21}v_1 + a_{22}v_2 + a_{23}v_3 + a_{24}v_4 \\ &= 0 \times 1 + 2 \times 2 + 1 \times 0 + (-2) \times 1 = 2.\end{aligned}$$

You should do the rest of the problem and verify

$$\begin{pmatrix} 1 & 2 & 1 & 3 \\ 0 & 2 & 1 & -2 \\ 2 & 1 & 4 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 8 \\ 2 \\ 5 \end{pmatrix}.$$

With this done, the next task is to multiply an $m \times n$ matrix times an $n \times p$ matrix. Before doing so, the following may be helpful.

$$(m \times \overbrace{n)\ (n}^{\text{these must match}} \times p \quad) = m \times p$$

If the two middle numbers don't match, you can't multiply the matrices!

Definition 2.1.5 *Let A be an $m \times n$ matrix and let B be an $n \times p$ matrix. Then B is of the form*

$$B = (\mathbf{b}_1, \cdots, \mathbf{b}_p)$$

where $\mathbf{b}_k$ is an $n \times 1$ matrix. Then an $m \times p$ matrix AB is defined as follows:

$$AB \equiv (A\mathbf{b}_1, \cdots, A\mathbf{b}_p) \tag{2.10}$$

where $A\mathbf{b}_k$ is an $m \times 1$ matrix. Hence AB as just defined is an $m \times p$ matrix. For example,

Example 2.1.6 *Multiply the following.*

$$\begin{pmatrix} 1 & 2 & 1 \\ 0 & 2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 0 \\ 0 & 3 & 1 \\ -2 & 1 & 1 \end{pmatrix}$$

The first thing you need to check before doing anything else is whether it is possible to do the multiplication. The first matrix is a 2×3 and the second matrix is a 3×3. Therefore, is it possible to multiply these matrices. According to the above discussion it should be a 2×3 matrix of the form

$$\left(\overbrace{\begin{pmatrix} 1 & 2 & 1 \\ 0 & 2 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ -2 \end{pmatrix}}^{\text{First column}}, \overbrace{\begin{pmatrix} 1 & 2 & 1 \\ 0 & 2 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix}}^{\text{Second column}}, \overbrace{\begin{pmatrix} 1 & 2 & 1 \\ 0 & 2 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}}^{\text{Third column}} \right)$$

You know how to multiply a matrix times a vector and so you do so to obtain each of the three columns. Thus

$$\begin{pmatrix} 1 & 2 & 1 \\ 0 & 2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 0 \\ 0 & 3 & 1 \\ -2 & 1 & 1 \end{pmatrix} = \begin{pmatrix} -1 & 9 & 3 \\ -2 & 7 & 3 \end{pmatrix}.$$

Here is another example.

Example 2.1.7 *Multiply the following.*

$$\begin{pmatrix} 1 & 2 & 0 \\ 0 & 3 & 1 \\ -2 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 1 \\ 0 & 2 & 1 \end{pmatrix}$$

First check if it is possible. This is of the form $(3 \times 3)(2 \times 3)$. The inside numbers do not match and so you can't do this multiplication. This means that anything you write will be absolute nonsense because it is impossible to multiply these matrices in this order. Aren't they the same two matrices considered in the previous example? Yes they are. It is just that here they are in a different order. This shows something you must always remember about matrix multiplication.

$$\boxed{\textbf{Order Matters!}}$$

Matrix multiplication is not commutative. This is very different than multiplication of numbers!

2.1.1 The ij^{th} Entry Of A Product

It is important to describe matrix multiplication in terms of entries of the matrices. What is the ij^{th} entry of AB? It would be the i^{th} entry of the j^{th} column of AB. Thus it would be the i^{th} entry of $A\mathbf{b}_j$. Now

$$\mathbf{b}_j = \begin{pmatrix} B_{1j} \\ \vdots \\ B_{nj} \end{pmatrix}$$

and from the above definition, the i^{th} entry is

$$\sum_{k=1}^{n} A_{ik} B_{kj}. \tag{2.11}$$

In terms of pictures of the matrix, you are doing

$$\begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{pmatrix} \begin{pmatrix} B_{11} & B_{12} & \cdots & B_{1p} \\ B_{21} & B_{22} & \cdots & B_{2p} \\ \vdots & \vdots & & \vdots \\ B_{n1} & B_{n2} & \cdots & B_{np} \end{pmatrix}$$

Then as explained above, the j^{th} column is of the form

$$\begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{pmatrix} \begin{pmatrix} B_{1j} \\ B_{2j} \\ \vdots \\ B_{nj} \end{pmatrix}$$

which is a $m \times 1$ matrix or column vector which equals

$$\begin{pmatrix} A_{11} \\ A_{21} \\ \vdots \\ A_{m1} \end{pmatrix} B_{1j} + \begin{pmatrix} A_{12} \\ A_{22} \\ \vdots \\ A_{m2} \end{pmatrix} B_{2j} + \cdots + \begin{pmatrix} A_{1n} \\ A_{2n} \\ \vdots \\ A_{mn} \end{pmatrix} B_{nj}.$$

The i^{th} entry of this $m \times 1$ matrix is

$$A_{i1}B_{1j} + A_{i2}B_{2j} + \cdots + A_{in}B_{nj} = \sum_{k=1}^{m} A_{ik}B_{kj}.$$

This shows the following definition for matrix multiplication in terms of the ij^{th} entries of the product harmonizes with Definition 2.1.3.

This motivates the definition for matrix multiplication which identifies the ij^{th} entries of the product.

Definition 2.1.8 *Let $A = (A_{ij})$ be an $m \times n$ matrix and let $B = (B_{ij})$ be an $n \times p$ matrix. Then AB is an $m \times p$ matrix and*

$$(AB)_{ij} = \sum_{k=1}^{n} A_{ik}B_{kj}. \tag{2.12}$$

Two matrices, A and B are said to be conformable in a particular order if they can be multiplied in that order. Thus if A is an $r \times s$ matrix and B is a $s \times p$ then A and B are conformable in the order AB. The above formula for $(AB)_{ij}$ says that it equals the i^{th} row of A times the j^{th} column of B.

Example 2.1.9 *Multiply if possible* $\begin{pmatrix} 1 & 2 \\ 3 & 1 \\ 2 & 6 \end{pmatrix} \begin{pmatrix} 2 & 3 & 1 \\ 7 & 6 & 2 \end{pmatrix}$.

First check to see if this is possible. It is of the form $(3 \times 2)(2 \times 3)$ and since the inside numbers match, it must be possible to do this and the result should be a 3×3 matrix. The answer is of the form

$$\left(\begin{pmatrix} 1 & 2 \\ 3 & 1 \\ 2 & 6 \end{pmatrix} \begin{pmatrix} 2 \\ 7 \end{pmatrix}, \begin{pmatrix} 1 & 2 \\ 3 & 1 \\ 2 & 6 \end{pmatrix} \begin{pmatrix} 3 \\ 6 \end{pmatrix}, \begin{pmatrix} 1 & 2 \\ 3 & 1 \\ 2 & 6 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right)$$

where the commas separate the columns in the resulting product. Thus the above product equals

$$\begin{pmatrix} 16 & 15 & 5 \\ 13 & 15 & 5 \\ 46 & 42 & 14 \end{pmatrix},$$

a 3×3 matrix as desired. In terms of the ij^{th} entries and the above definition, the entry in the third row and second column of the product should equal

$$\sum_j a_{3k} b_{k2} = a_{31} b_{12} + a_{32} b_{22} = 2 \times 3 + 6 \times 6 = 42.$$

You should try a few more such examples to verify the above definition in terms of the ij^{th} entries works for other entries.

Example 2.1.10 *Multiply if possible* $\begin{pmatrix} 1 & 2 \\ 3 & 1 \\ 2 & 6 \end{pmatrix} \begin{pmatrix} 2 & 3 & 1 \\ 7 & 6 & 2 \\ 0 & 0 & 0 \end{pmatrix}$.

This is not possible because it is of the form $(3 \times 2)(3 \times 3)$ and the middle numbers don't match.

Example 2.1.11 *Multiply if possible* $\begin{pmatrix} 2 & 3 & 1 \\ 7 & 6 & 2 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 1 \\ 2 & 6 \end{pmatrix}$.

This is possible because in this case it is of the form $(3 \times 3)(3 \times 2)$ and the middle numbers do match. When the multiplication is done it equals

$$\begin{pmatrix} 13 & 13 \\ 29 & 32 \\ 0 & 0 \end{pmatrix}.$$

Check this and be sure you come up with the same answer.

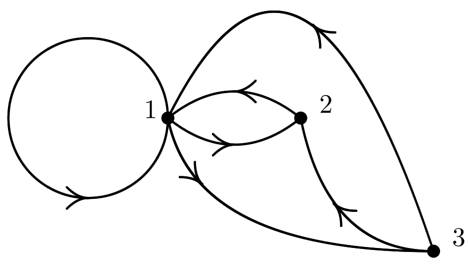
Example 2.1.12 *Multiply if possible* $\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 1 & 0 \end{pmatrix}$.

In this case you are trying to do $(3 \times 1)(1 \times 4)$. The inside numbers match so you can do it. Verify

$$\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 1 & 0 \\ 2 & 4 & 2 & 0 \\ 1 & 2 & 1 & 0 \end{pmatrix}$$

2.1.2 Digraphs

Consider the following graph illustrated in the picture.



There are three locations in this graph, labelled 1,2, and 3. The directed lines represent a way of going from one location to another. Thus there is one way to go from location 1 to location 1. There is one way to go from location 1 to location 3. It is not possible to go from location 2 to location 3 although it is possible to go from location 3 to location 2. Lets refer to moving along one of these directed lines as a step. The following 3×3 matrix is a numerical way of writing the above graph. This is sometimes called a digraph, short for directed graph.

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix}$$

Thus a_{ij}, the entry in the i^{th} row and j^{th} column represents the number of ways to go from location i to location j in one step.

Problem: Find the number of ways to go from i to j using exactly k steps.

Denote the answer to the above problem by a_{ij}^k. We don't know what it is right now unless $k = 1$ when it equals a_{ij} described above. However, if we did know what it was, we could find a_{ij}^{k+1} as follows.

$$a_{ij}^{k+1} = \sum_r a_{ir}^k a_{rj}$$

This is because if you go from i to j in $k+1$ steps, you first go from i to r in k steps and then for each of these ways there are a_{rj} ways to go from there to j. Thus $a_{ir}^k a_{rj}$ gives the number of ways to go from i to j in $k+1$ steps such that the k^{th} step leaves you at location r. Adding these gives the above sum. Now you recognize this as the ij^{th} entry of the product of two matrices. Thus

$$a_{ij}^2 = \sum_r a_{ir} a_{rj}, \quad a_{ij}^3 = \sum_r a_{ir}^2 a_{rj}$$

and so forth. From the above definition of matrix multiplication, this shows that if A is the matrix associated with the directed graph as above, then a_{ij}^k is just the ij^{th} entry of A^k where A^k is just what you would think it should be, A multiplied by itself k times.

Thus in the above example, to find the number of ways of going from 1 to 3 in two steps you would take that matrix and multiply it by itself and then take the entry in the first row and third column. Thus

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix}^2 = \begin{pmatrix} 3 & 2 & 1 \\ 1 & 1 & 1 \\ 2 & 1 & 1 \end{pmatrix}$$

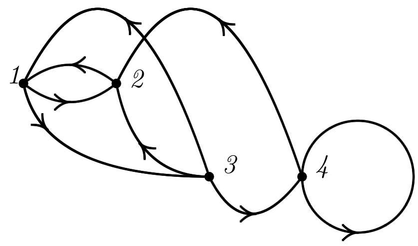
and you see there is exactly one way to go from 1 to 3 in two steps. You can easily see this is true from looking at the graph also. Note there are three ways to go from 1 to 1 in 2 steps. Can you find them from the graph? What would you do if you wanted to consider 5 steps?

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix}^5 = \begin{pmatrix} 28 & 19 & 13 \\ 13 & 9 & 6 \\ 19 & 13 & 9 \end{pmatrix}$$

There are 19 ways to go from 1 to 2 in five steps. Do you think you could list them all by looking at the graph? I don't think you could do it without wasting a lot of time.

Of course there is nothing sacred about having only three locations. Everything works just as well with any number of locations. In general if you have n locations, you would need to use a $n \times n$ matrix.

Example 2.1.13 *Consider the following directed graph.*



Write the matrix which is associated with this directed graph and find the number of ways to go from 2 to 4 in three steps.

Here you need to use a 4×4 matrix. The one you need is

$$\begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \end{pmatrix}$$

Then to find the answer, you just need to multiply this matrix by itself three times and look at the entry in the second row and fourth column.

$$\begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \end{pmatrix}^3 = \begin{pmatrix} 1 & 3 & 2 & 1 \\ 2 & 1 & 0 & 1 \\ 3 & 3 & 1 & 2 \\ 1 & 2 & 1 & 1 \end{pmatrix}$$

There is exactly one way to go from 2 to 4 in three steps.

How many ways would there be of going from 2 to 4 in five steps?

$$\begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \end{pmatrix}^5 = \begin{pmatrix} 5 & 9 & 5 & 4 \\ 5 & 4 & 1 & 3 \\ 9 & 10 & 4 & 6 \\ 4 & 6 & 3 & 3 \end{pmatrix}$$

There are three ways. Note there are 10 ways to go from 3 to 2 in five steps.

This is an interesting application of the concept of the ij^{th} entry of the product matrices.

2.1.3 Properties Of Matrix Multiplication

As pointed out above, sometimes it is possible to multiply matrices in one order but not in the other order. What if it makes sense to multiply them in either order? Will they be equal then?

Example 2.1.14 *Compare* $\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ *and* $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$.

The first product is

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 4 & 3 \end{pmatrix},$$

the second product is

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 3 & 4 \\ 1 & 2 \end{pmatrix},$$

and you see these are not equal. Therefore, you cannot conclude that $AB = BA$ for matrix multiplication. However, there are some properties which do hold.

Proposition 2.1.15 *If all multiplications and additions make sense, the following hold for matrices, A, B, C and a, b scalars.*

$$A(aB + bC) = a(AB) + b(AC) \tag{2.13}$$

$$(B + C)A = BA + CA \tag{2.14}$$

$$A(BC) = (AB)C \tag{2.15}$$

Proof: Using the above definition of matrix multiplication,

$$\begin{aligned} (A(aB + bC))_{ij} &= \sum_k A_{ik}(aB + bC)_{kj} \\ &= \sum_k A_{ik}(aB_{kj} + bC_{kj}) \\ &= a\sum_k A_{ik}B_{kj} + b\sum_k A_{ik}C_{kj} \\ &= a(AB)_{ij} + b(AC)_{ij} \\ &= (a(AB) + b(AC))_{ij} \end{aligned}$$

showing that $A(B + C) = AB + AC$ as claimed. Formula 2.14 is entirely similar.

Consider 2.15, the associative law of multiplication. Before reading this, review the definition of matrix multiplication in terms of entries of the matrices.

$$\begin{aligned} (A(BC))_{ij} &= \sum_k A_{ik}(BC)_{kj} \\ &= \sum_k A_{ik}\sum_l B_{kl}C_{lj} \\ &= \sum_l (AB)_{il}C_{lj} \\ &= ((AB)C)_{ij}. \blacksquare \end{aligned}$$

Another important operation on matrices is that of taking the transpose. The following example shows what is meant by this operation, denoted by placing a T as an exponent on the matrix.

$$\begin{pmatrix} 1 & 1+2i \\ 3 & 1 \\ 2 & 6 \end{pmatrix}^T = \begin{pmatrix} 1 & 3 & 2 \\ 1+2i & 1 & 6 \end{pmatrix}$$

What happened? The first column became the first row and the second column became the second row. Thus the 3×2 matrix became a 2×3 matrix. The number 3 was in the second row and the first column and it ended up in the first row and second column. This motivates the following definition of the transpose of a matrix.

Definition 2.1.16 *Let A be an $m \times n$ matrix. Then A^T denotes the $n \times m$ matrix which is defined as follows.*

$$\left(A^T\right)_{ij} = A_{ji}$$

The transpose of a matrix has the following important property.

Lemma 2.1.17 *Let A be an $m \times n$ matrix and let B be a $n \times p$ matrix. Then*

$$(AB)^T = B^T A^T \tag{2.16}$$

and if α and β are scalars,

$$(\alpha A + \beta B)^T = \alpha A^T + \beta B^T \tag{2.17}$$

Proof: From the definition,

$$\begin{aligned}
\left((AB)^T\right)_{ij} &= (AB)_{ji} \\
&= \sum_k A_{jk} B_{ki} \\
&= \sum_k \left(B^T\right)_{ik} \left(A^T\right)_{kj} \\
&= \left(B^T A^T\right)_{ij}
\end{aligned}$$

2.17 is left as an exercise. ■

Definition 2.1.18 *An $n \times n$ matrix A is said to be symmetric if $A = A^T$. It is said to be skew symmetric if $A^T = -A$.*

Example 2.1.19 *Let*

$$A = \begin{pmatrix} 2 & 1 & 3 \\ 1 & 5 & -3 \\ 3 & -3 & 7 \end{pmatrix}.$$

Then A is symmetric.

Example 2.1.20 *Let*

$$A = \begin{pmatrix} 0 & 1 & 3 \\ -1 & 0 & 2 \\ -3 & -2 & 0 \end{pmatrix}$$

Then A is skew symmetric.

There is a special matrix called I and defined by

$$I_{ij} = \delta_{ij}$$

where δ_{ij} is the Kronecker symbol defined by

$$\delta_{ij} = \begin{cases} 1 \text{ if } i = j \\ 0 \text{ if } i \neq j \end{cases}$$

It is called the identity matrix because it is a multiplicative identity in the following sense.

Lemma 2.1.21 *Suppose A is an $m \times n$ matrix and I_n is the $n \times n$ identity matrix. Then $AI_n = A$. If I_m is the $m \times m$ identity matrix, it also follows that $I_m A = A$.*

Proof:

$$\begin{aligned}
(AI_n)_{ij} &= \sum_k A_{ik} \delta_{kj} \\
&= A_{ij}
\end{aligned}$$

and so $AI_n = A$. The other case is left as an exercise for you.

Definition 2.1.22 *An $n \times n$ matrix A has an inverse A^{-1} if and only if there exists a matrix, denoted as A^{-1} such that $AA^{-1} = A^{-1}A = I$ where $I = (\delta_{ij})$ for*

$$\delta_{ij} \equiv \begin{cases} 1 \text{ if } i = j \\ 0 \text{ if } i \neq j \end{cases}$$

Such a matrix is called invertible.

If it acts like an inverse, then it is the inverse. This is the message of the following proposition.

Proposition 2.1.23 *Suppose $AB = BA = I$. Then $B = A^{-1}$.*

Proof: From the definition B is an inverse for A. Could there be another one B'?

$$B' = B'I = B'(AB) = (B'A)B = IB = B.$$

Thus, the inverse, if it exists, is unique. ■

2.1.4 Finding The Inverse Of A Matrix

A little later a formula is given for the inverse of a matrix. However, it is not a good way to find the inverse for a matrix. There is a much easier way and it is this which is presented here. It is also important to note that not all matrices have inverses.

Example 2.1.24 *Let* $A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$. *Does* A *have an inverse?*

One might think A would have an inverse because it does not equal zero. However,

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

and if A^{-1} existed, this could not happen because you could multiply on the left by the inverse A and conclude the vector $(-1, 1)^T = (0, 0)^T$. Thus the answer is that A does not have an inverse.

Suppose you want to find B such that $AB = I$. Let

$$B = \begin{pmatrix} \mathbf{b}_1 & \cdots & \mathbf{b}_n \end{pmatrix}$$

Also the i^{th} column of I is

$$\mathbf{e}_i = \begin{pmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{pmatrix}^T$$

Thus, if $AB = I$, $\mathbf{b}_i$, the i^{th} column of B must satisfy the equation $A\mathbf{b}_i = \mathbf{e}_i$. The augmented matrix for finding $\mathbf{b}_i$ is $(A|\mathbf{e}_i)$. Thus, by doing row operations till A becomes I, you end up with $(I|\mathbf{b}_i)$ where $\mathbf{b}_i$ is the solution to $A\mathbf{b}_i = \mathbf{e}_i$. Now the same sequence of row operations works regardless of the right side of the agumented matrix $(A|\mathbf{e}_i)$ and so you can save trouble by simply doing the following.

$$(A|I) \overset{\text{row operations}}{\rightarrow} (I|B)$$

and the i^{th} column of B is $\mathbf{b}_i$, the solution to $A\mathbf{b}_i = \mathbf{e}_i$. Thus $AB = I$.

This is the reason for the following simple procedure for finding the inverse of a matrix. This procedure is called the Gauss Jordan procedure. It produces the inverse if the matrix has one. Actually, it produces the right inverse.

Procedure 2.1.25 *Suppose* A *is an* $n \times n$ *matrix. To find* A^{-1} *if it exists, form the augmented* $n \times 2n$ *matrix,*

$$(A|I)$$

and then do row operations until you obtain an $n \times 2n$ *matrix of the form*

$$(I|B) \tag{2.18}$$

if possible. When this has been done, $B = A^{-1}$. *The matrix* A *has an inverse exactly when it is possible to do row operations and end up with one like 2.18.*

As described above, the following is a description of what you have just done.

$$A \overset{R_q R_{q-1} \cdots R_1}{\rightarrow} I$$
$$I \overset{R_q R_{q-1} \cdots R_1}{\rightarrow} B$$

where those R_i sympolize row operations. It follows that you could undo what you did by doing the inverse of these row operations in the opposite order. Thus

$$I \overset{R_1^{-1}\cdots R_{q-1}^{-1}R_q^{-1}}{\rightarrow} A$$
$$B \overset{R_1^{-1}\cdots R_{q-1}^{-1}R_q^{-1}}{\rightarrow} I$$

Here R^{-1} is the row operation which undoes the row operation R. Therefore, if you form $(B|I)$ and do the inverse of the row operations which produced I from A in the reverse order, you would obtain $(I|A)$. By the same reasoning above, it follows that A is a right inverse of B and so $BA = I$ also. It follows from Proposition 2.1.23 that $B = A^{-1}$. Thus the procedure produces **the** inverse whenever it works.

If it is possible to do row operations and end up with $A \overset{\text{row operations}}{\rightarrow} I$, then the above argument shows that A has an inverse. Conversely, if A has an inverse, can it be found by the above procedure? In this case there exists a unique solution $\mathbf{x}$ to the equation $A\mathbf{x} = \mathbf{y}$. In fact it is just $\mathbf{x} = I\mathbf{x} = A^{-1}\mathbf{y}$. Thus in terms of augmented matrices, you would expect to obtain

$$(A|\mathbf{y}) \rightarrow \left(I|A^{-1}\mathbf{y}\right)$$

That is, you would expect to be able to do row operations to A and end up with I.

The details will be explained fully when a more careful discussion is given which is based on more fundamental considerations. For now, it suffices to observe that whenever the above procedure works, it finds the inverse.

Example 2.1.26 *Let* $A = \begin{pmatrix} 1 & 0 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \end{pmatrix}$. *Find* A^{-1}.

Form the augmented matrix

$$\begin{pmatrix} 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & -1 & 1 & 0 & 1 & 0 \\ 1 & 1 & -1 & 0 & 0 & 1 \end{pmatrix}.$$

Now do row operations until the $n \times n$ matrix on the left becomes the identity matrix. This yields after some computations,

$$\begin{pmatrix} 1 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 1 & 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 1 & -\frac{1}{2} & -\frac{1}{2} \end{pmatrix}$$

and so the inverse of A is the matrix on the right,

$$\begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & -1 & 0 \\ 1 & -\frac{1}{2} & -\frac{1}{2} \end{pmatrix}.$$

Checking the answer is easy. Just multiply the matrices and see if it works.

$$\begin{pmatrix} 1 & 0 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \end{pmatrix} \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & -1 & 0 \\ 1 & -\frac{1}{2} & -\frac{1}{2} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Always check your answer because if you are like some of us, you will usually have made a mistake.

Example 2.1.27 *Let* $A = \begin{pmatrix} 1 & 2 & 2 \\ 1 & 0 & 2 \\ 3 & 1 & -1 \end{pmatrix}$. *Find* A^{-1}.

Set up the augmented matrix $(A|I)$

$$\begin{pmatrix} 1 & 2 & 2 & 1 & 0 & 0 \\ 1 & 0 & 2 & 0 & 1 & 0 \\ 3 & 1 & -1 & 0 & 0 & 1 \end{pmatrix}$$

Next take (-1) times the first row and add to the second followed by (-3) times the first row added to the last. This yields

$$\begin{pmatrix} 1 & 2 & 2 & 1 & 0 & 0 \\ 0 & -2 & 0 & -1 & 1 & 0 \\ 0 & -5 & -7 & -3 & 0 & 1 \end{pmatrix}.$$

Then take 5 times the second row and add to -2 times the last row.

$$\begin{pmatrix} 1 & 2 & 2 & 1 & 0 & 0 \\ 0 & -10 & 0 & -5 & 5 & 0 \\ 0 & 0 & 14 & 1 & 5 & -2 \end{pmatrix}$$

Next take the last row and add to (-7) times the top row. This yields

$$\begin{pmatrix} -7 & -14 & 0 & -6 & 5 & -2 \\ 0 & -10 & 0 & -5 & 5 & 0 \\ 0 & 0 & 14 & 1 & 5 & -2 \end{pmatrix}.$$

Now take $(-7/5)$ times the second row and add to the top.

$$\begin{pmatrix} -7 & 0 & 0 & 1 & -2 & -2 \\ 0 & -10 & 0 & -5 & 5 & 0 \\ 0 & 0 & 14 & 1 & 5 & -2 \end{pmatrix}.$$

Finally divide the top row by -7, the second row by -10 and the bottom row by 14 which yields

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 0 & -\frac{1}{7} & \frac{2}{7} & \frac{2}{7} \\ 0 & 1 & 0 & \frac{1}{2} & -\frac{1}{2} & 0 \\ 0 & 0 & 1 & \frac{1}{14} & \frac{5}{14} & -\frac{1}{7} \end{array}\right).$$

Therefore, the inverse is

$$\begin{pmatrix} -\frac{1}{7} & \frac{2}{7} & \frac{2}{7} \\ \frac{1}{2} & -\frac{1}{2} & 0 \\ \frac{1}{14} & \frac{5}{14} & -\frac{1}{7} \end{pmatrix}$$

Example 2.1.28 *Let* $A = \begin{pmatrix} 1 & 2 & 2 \\ 1 & 0 & 2 \\ 2 & 2 & 4 \end{pmatrix}$. *Find* A^{-1}.

Write the augmented matrix $(A|I)$

$$\left(\begin{array}{cccccc} 1 & 2 & 2 & 1 & 0 & 0 \\ 1 & 0 & 2 & 0 & 1 & 0 \\ 2 & 2 & 4 & 0 & 0 & 1 \end{array}\right)$$

and proceed to do row operations attempting to obtain $\left(I|A^{-1}\right)$. Take (-1) times the top row and add to the second. Then take (-2) times the top row and add to the bottom.

$$\left(\begin{array}{cccccc} 1 & 2 & 2 & 1 & 0 & 0 \\ 0 & -2 & 0 & -1 & 1 & 0 \\ 0 & -2 & 0 & -2 & 0 & 1 \end{array}\right)$$

Next add (-1) times the second row to the bottom row.

$$\left(\begin{array}{cccccc} 1 & 2 & 2 & 1 & 0 & 0 \\ 0 & -2 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & -1 & 1 \end{array}\right)$$

At this point, you can see there will be no inverse because you have obtained a row of zeros in the left half of the augmented matrix $(A|I)$. Thus there will be no way to obtain I on the left. In other words, the three systems of equations you must solve to find the inverse have no solution. In particular, there is no solution for the first column of A^{-1} which must solve

$$A\left(\begin{array}{c} x \\ y \\ z \end{array}\right) = \left(\begin{array}{c} 1 \\ 0 \\ 0 \end{array}\right)$$

because a sequence of row operations leads to the impossible equation, $0x + 0y + 0z = -1$.

2.2 Exercises

1. In 2.1 - 2.8 describe $-A$ and 0.
2. Let A be an $n \times n$ matrix. Show A equals the sum of a symmetric and a skew symmetric matrix.
3. Show every skew symmetric matrix has all zeros down the main diagonal. The main diagonal consists of every entry of the matrix which is of the form a_{ii}. It runs from the upper left down to the lower right.
4. Using only the properties 2.1 - 2.8 show $-A$ is unique.
5. Using only the properties 2.1 - 2.8 show 0 is unique.
6. Using only the properties 2.1 - 2.8 show $0A = 0$. Here the 0 on the left is the scalar 0 and the 0 on the right is the zero for $m \times n$ matrices.
7. Using only the properties 2.1 - 2.8 and previous problems show $(-1)\,A = -A$.
8. Prove 2.17.
9. Prove that $I_m A = A$ where A is an $m \times n$ matrix.
10. Let A and be a real $m \times n$ matrix and let $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^m$. Show $(A\mathbf{x}, \mathbf{y})_{\mathbb{R}^m} = \left(\mathbf{x}, A^T\mathbf{y}\right)_{\mathbb{R}^n}$ where $(\cdot,\cdot)_{\mathbb{R}^k}$ denotes the dot product in $\mathbb{R}^k$.

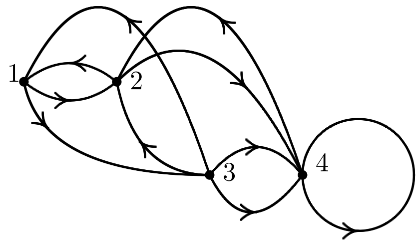
11. Use the result of Problem 10 to verify directly that $(AB)^T = B^T A^T$ without making any reference to subscripts.

12. Let $\mathbf{x} = (-1,-1,1)$ and $\mathbf{y} = (0,1,2)$. Find $\mathbf{x}^T\mathbf{y}$ and $\mathbf{x}\mathbf{y}^T$ if possible.

13. Give an example of matrices, A, B, C such that $B \neq C$, $A \neq 0$, and yet $AB = AC$.

14. Let $A = \begin{pmatrix} 1 & 1 \\ -2 & -1 \\ 1 & 2 \end{pmatrix}$, $B = \begin{pmatrix} 1 & -1 & -2 \\ 2 & 1 & -2 \end{pmatrix}$, and $C = \begin{pmatrix} 1 & 1 & -3 \\ -1 & 2 & 0 \\ -3 & -1 & 0 \end{pmatrix}$. Find if possible the following products. AB, BA, AC, CA, CB, BC.

15. Consider the following digraph.



Write the matrix associated with this digraph and find the number of ways to go from 3 to 4 in three steps.

16. Show that if A^{-1} exists for an $n \times n$ matrix, then it is unique. That is, if $BA = I$ and $AB = I$, then $B = A^{-1}$.

17. Show $(AB)^{-1} = B^{-1}A^{-1}$.

18. Show that if A is an invertible $n \times n$ matrix, then so is A^T and $\left(A^T\right)^{-1} = \left(A^{-1}\right)^T$.

19. Show that if A is an $n \times n$ invertible matrix and $\mathbf{x}$ is a $n \times 1$ matrix such that $A\mathbf{x} = \mathbf{b}$ for $\mathbf{b}$ an $n \times 1$ matrix, then $\mathbf{x} = A^{-1}\mathbf{b}$.

20. Give an example of a matrix A such that $A^2 = I$ and yet $A \neq I$ and $A \neq -I$.

21. Give an example of matrices, A, B such that neither A nor B equals zero and yet $AB = 0$.

22. Write $\begin{pmatrix} x_1 - x_2 + 2x_3 \\ 2x_3 + x_1 \\ 3x_3 \\ 3x_4 + 3x_2 + x_1 \end{pmatrix}$ in the form $A\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}$ where A is an appropriate matrix.

23. Give another example other than the one given in this section of two square matrices, A and B such that $AB \neq BA$.

24. Suppose A and B are square matrices of the same size. Which of the following are correct?

    (a) $(A-B)^2 = A^2 - 2AB + B^2$
    (b) $(AB)^2 = A^2B^2$
    (c) $(A+B)^2 = A^2 + 2AB + B^2$
    (d) $(A+B)^2 = A^2 + AB + BA + B^2$

(e) $A^2B^2 = A(AB)B$

(f) $(A+B)^3 = A^3 + 3A^2B + 3AB^2 + B^3$

(g) $(A+B)(A-B) = A^2 - B^2$

(h) None of the above. They are all wrong.

(i) All of the above. They are all right.

25. Let $A = \begin{pmatrix} -1 & -1 \\ 3 & 3 \end{pmatrix}$. Find all 2×2 matrices, B such that $AB = 0$.

26. Prove that if A^{-1} exists and $A\mathbf{x} = \mathbf{0}$ then $\mathbf{x} = \mathbf{0}$.

27. Let
$$A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 4 \\ 1 & 0 & 2 \end{pmatrix}.$$
Find A^{-1} if possible. If A^{-1} does not exist, determine why.

28. Let
$$A = \begin{pmatrix} 1 & 0 & 3 \\ 2 & 3 & 4 \\ 1 & 0 & 2 \end{pmatrix}.$$
Find A^{-1} if possible. If A^{-1} does not exist, determine why.

29. Let
$$A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 4 \\ 4 & 5 & 10 \end{pmatrix}.$$
Find A^{-1} if possible. If A^{-1} does not exist, determine why.

30. Let
$$A = \begin{pmatrix} 1 & 2 & 0 & 2 \\ 1 & 1 & 2 & 0 \\ 2 & 1 & -3 & 2 \\ 1 & 2 & 1 & 2 \end{pmatrix}$$
Find A^{-1} if possible. If A^{-1} does not exist, determine why.

2.3 Linear Transformations

By 2.13, if A is an $m \times n$ matrix, then for $\mathbf{v}, \mathbf{u}$ vectors in $\mathbb{F}^n$ and a, b scalars,

$$A\left(\overbrace{a\mathbf{u}+b\mathbf{v}}^{\in\mathbb{F}^n}\right) = aA\mathbf{u} + bA\mathbf{v} \in \mathbb{F}^m \tag{2.19}$$

Definition 2.3.1 *A function, $A : \mathbb{F}^n \to \mathbb{F}^m$ is called a linear transformation if for all $\mathbf{u}, \mathbf{v} \in \mathbb{F}^n$ and a, b scalars, 2.19 holds.*

From 2.19, matrix multiplication defines a linear transformation as just defined. It turns out this is the only type of linear transformation available. Thus if A is a linear transformation from $\mathbb{F}^n$ to $\mathbb{F}^m$, there is always a matrix which produces A. Before showing this, here is a simple definition.

Definition 2.3.2 *A vector,* $\mathbf{e}_i \in \mathbb{F}^n$ *is defined as follows:*

$$\mathbf{e}_i \equiv \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix},$$

where the 1 is in the i^{th} *position and there are zeros everywhere else. Thus*

$$\mathbf{e}_i = (0, \cdots, 0, 1, 0, \cdots, 0)^T.$$

Of course the $\mathbf{e}_i$ for a particular value of i in $\mathbb{F}^n$ would be different than the $\mathbf{e}_i$ for that same value of i in $\mathbb{F}^m$ for $m \neq n$. One of them is longer than the other. However, which one is meant will be determined by the context in which they occur.

These vectors have a significant property.

Lemma 2.3.3 *Let* $\mathbf{v} \in \mathbb{F}^n$. *Thus* $\mathbf{v}$ *is a list of numbers arranged vertically,* $v_1, \cdots, v_n$. *Then*

$$\mathbf{e}_i^T \mathbf{v} = v_i. \tag{2.20}$$

Also, if A *is an* $m \times n$ *matrix, then letting* $\mathbf{e}_i \in \mathbb{F}^m$ *and* $\mathbf{e}_j \in \mathbb{F}^n$,

$$\mathbf{e}_i^T A \mathbf{e}_j = A_{ij} \tag{2.21}$$

Proof: First note that $\mathbf{e}_i^T$ is a $1 \times n$ matrix and $\mathbf{v}$ is an $n \times 1$ matrix so the above multiplication in 2.20 makes perfect sense. It equals

$$(0, \cdots, 1, \cdots 0) \begin{pmatrix} v_1 \\ \vdots \\ v_i \\ \vdots \\ v_n \end{pmatrix} = v_i$$

as claimed.

Consider 2.21. From the definition of matrix multiplication, and noting that $(\mathbf{e}_j)_k = \delta_{kj}$

$$\mathbf{e}_i^T A \mathbf{e}_j = \mathbf{e}_i^T \begin{pmatrix} \sum_k A_{1k} (\mathbf{e}_j)_k \\ \vdots \\ \sum_k A_{ik} (\mathbf{e}_j)_k \\ \vdots \\ \sum_k A_{mk} (\mathbf{e}_j)_k \end{pmatrix} = \mathbf{e}_i^T \begin{pmatrix} A_{1j} \\ \vdots \\ A_{ij} \\ \vdots \\ A_{mj} \end{pmatrix} = A_{ij}$$

by the first part of the lemma. ■

Theorem 2.3.4 *Let* $L : \mathbb{F}^n \to \mathbb{F}^m$ *be a linear transformation. Then there exists a unique* $m \times n$ *matrix* A *such that*

$$A\mathbf{x} = L\mathbf{x}$$

for all $\mathbf{x} \in \mathbb{F}^n$. *The* ik^{th} *entry of this matrix is given by*

$$\mathbf{e}_i^T L \mathbf{e}_k \tag{2.22}$$

Stated in another way, the k^{th} *column of* A *equals* $L\mathbf{e}_\mathbf{k}$.

Proof: By the lemma,

$$(L\mathbf{x})_i = \mathbf{e}_i^T L\mathbf{x} = \mathbf{e}_i^T x_k L\mathbf{e}_k = \left(\mathbf{e}_i^T L\mathbf{e}_k\right) x_k.$$

Let $A_{ik} = \mathbf{e}_i^T L\mathbf{e}_k$, to prove the existence part of the theorem.

To verify uniqueness, suppose $B\mathbf{x} = A\mathbf{x} = L\mathbf{x}$ for all $\mathbf{x} \in \mathbb{F}^n$. Then in particular, this is true for $\mathbf{x} = \mathbf{e}_j$ and then multiply on the left by $\mathbf{e}_i^T$ to obtain

$$B_{ij} = \mathbf{e}_i^T B\mathbf{e}_j = \mathbf{e}_i^T A\mathbf{e}_j = A_{ij}$$

showing $A = B$. ■

Corollary 2.3.5 *A linear transformation, $L : \mathbb{F}^n \to \mathbb{F}^m$ is completely determined by the vectors $\{L\mathbf{e}_1, \cdots, L\mathbf{e}_n\}$.*

Proof: This follows immediately from the above theorem. The unique matrix determining the linear transformation which is given in 2.22 depends only on these vectors. ■

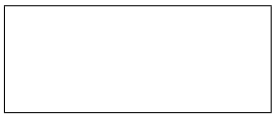
This theorem shows that any linear transformation defined on $\mathbb{F}^n$ can always be considered as a matrix. Therefore, the terms "linear transformation" and "matrix" are often used interchangeably. For example, to say that a matrix is one to one, means the linear transformation determined by the matrix is one to one.

Example 2.3.6 *Find the linear transformation, $L : \mathbb{R}^2 \to \mathbb{R}^2$ which has the property that $L\mathbf{e}_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ and $L\mathbf{e}_2 = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$. From the above theorem and corollary, this linear transformation is that determined by matrix multiplication by the matrix*

$$\begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix}.$$

Definition 2.3.7 *Let $L : \mathbb{F}^n \to \mathbb{F}^m$ be a linear transformation and let its matrix be the $m \times n$ matrix A. Then $\ker(L) \equiv \{\mathbf{x} \in \mathbb{F}^n : L\mathbf{x} = \mathbf{0}\}$. Sometimes people also write this as $N(A)$, the null space of A.*

Then there is a fundamental result in the case where $m < n$. In this case, the matrix A of the linear transformation looks like the following.



Theorem 2.3.8 *Let A be an $m \times n$ matrix where $m < n$. Then $N(A)$ contains nonzero vectors.*

Proof: First consider the case where A is a $1 \times n$ matrix for $n > 1$. Say

$$A = \begin{pmatrix} a_1 & \cdots & a_n \end{pmatrix}$$

If $a_1 = 0$, consider the vector $\mathbf{x} = \mathbf{e}_1$. If $a_1 \neq 0$, let

$$\mathbf{x} = \begin{pmatrix} b \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

where b is chosen to satisfy the equation

$$a_1 b + \sum_{k=2}^{n} a_k = 0$$

Suppose now that the theorem is true for any $m \times n$ matrix with $n > m$ and consider an $(m \times 1) \times n$ matrix A where $n > m + 1$. If the first column of A is $\mathbf{0}$, then you could let $\mathbf{x} = \mathbf{e}_1$ as above. If the first column is not the zero vector, then by doing row operations, the equation $A\mathbf{x} = \mathbf{0}$ can be reduced to the equivalent system

$$A_1 \mathbf{x} = \mathbf{0}$$

where A_1 is of the form

$$A_1 = \begin{pmatrix} 1 & \mathbf{a}^T \\ \mathbf{0} & B \end{pmatrix}$$

where B is an $m \times (n-1)$ matrix. Since $n > m + 1$, it follows that $(n-1) > m$ and so by induction, there exists a nonzero vector $\mathbf{y} \in \mathbb{F}^{n-1}$ such that $B\mathbf{y} = \mathbf{0}$. Then consider the vector

$$\mathbf{x} = \begin{pmatrix} b \\ \mathbf{y} \end{pmatrix}$$

$A_1\mathbf{x}$ has for its top entry the expression $b + \mathbf{a}^T\mathbf{y}$. Letting $B = \begin{pmatrix} \mathbf{b}_1^T \\ \vdots \\ \mathbf{b}_m^T \end{pmatrix}$, the i^{th} entry of $A_1\mathbf{x}$ for $i > 1$ is of the form $\mathbf{b}_i^T\mathbf{y} = 0$. Thus if b is chosen to satisfy the equation $b + \mathbf{a}^T\mathbf{y} = 0$, then $A_1\mathbf{x} = \mathbf{0}$.■

2.4 Subspaces And Spans

Definition 2.4.1 *Let $\{\mathbf{x}_1, \cdots, \mathbf{x}_p\}$ be vectors in $\mathbb{F}^n$. A linear combination is any expression of the form*

$$\sum_{i=1}^{p} c_i \mathbf{x}_i$$

where the c_i are scalars. The set of all linear combinations of these vectors is called $\operatorname{span}(\mathbf{x}_1, \cdots, \mathbf{x}_n)$. *If $V \subseteq \mathbb{F}^n$, then V is called a subspace if whenever α, β are scalars and $\mathbf{u}$ and $\mathbf{v}$ are vectors of V, it follows $\alpha\mathbf{u} + \beta\mathbf{v} \in V$. That is, it is "closed under the algebraic operations of vector addition and scalar multiplication". A linear combination of vectors is said to be trivial if all the scalars in the linear combination equal zero. A set of vectors is said to be linearly independent if the only linear combination of these vectors which equals the zero vector is the trivial linear combination. Thus $\{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$ is called* *linearly independent if whenever*

$$\sum_{k=1}^{p} c_k \mathbf{x}_k = \mathbf{0}$$

it follows that all the scalars c_k equal zero. A set of vectors, $\{\mathbf{x}_1, \cdots, \mathbf{x}_p\}$, is called linearly dependent if it is not linearly independent. Thus the set of vectors is linearly dependent if there exist scalars $c_i, i = 1, \cdots, n$, not all zero such that $\sum_{k=1}^{p} c_k \mathbf{x}_k = \mathbf{0}$.

Proposition 2.4.2 *Let $V \subseteq \mathbb{F}^n$. Then V is a subspace if and only if it is a vector space itself with respect to the same operations of scalar multiplication and vector addition.*

Proof: Suppose first that V is a subspace. All algebraic properties involving scalar multiplication and vector addition hold for V because these things hold for $\mathbb{F}^n$. Is $\mathbf{0} \in V$? Yes it is. This is because $0\mathbf{v} \in V$ and $0\mathbf{v} = \mathbf{0}$. By assumption, for α a scalar and $\mathbf{v} \in V, \alpha\mathbf{v} \in V$. Therefore, $-\mathbf{v} = (-1)\mathbf{v} \in V$. Thus V has the additive identity and additive inverse. By assumption, V is closed with respect to the two operations. Thus V is a vector space. If $V \subseteq \mathbb{F}^n$ is a vector space, then by definition, if α, β are scalars and $\mathbf{u}, \mathbf{v}$ vectors in V, it follows that $\alpha\mathbf{v} + \beta\mathbf{u} \in V$. ■

Thus, from the above, subspaces of $\mathbb{F}^n$ are just subsets of $\mathbb{F}^n$ which are themselves vector spaces.

Lemma 2.4.3 *A set of vectors $\{\mathbf{x}_1, \cdots, \mathbf{x}_p\}$ is linearly independent if and only if none of the vectors can be obtained as a linear combination of the others.*

Proof: Suppose first that $\{\mathbf{x}_1, \cdots, \mathbf{x}_p\}$ is linearly independent. If $\mathbf{x}_k = \sum_{j \neq k} c_j \mathbf{x}_j$, then

$$\mathbf{0} = 1\mathbf{x}_k + \sum_{j \neq k} (-c_j)\,\mathbf{x}_j,$$

a nontrivial linear combination, contrary to assumption. This shows that if the set is linearly independent, then none of the vectors is a linear combination of the others.

Now suppose no vector is a linear combination of the others. Is $\{\mathbf{x}_1, \cdots, \mathbf{x}_p\}$ linearly independent? If it is not, there exist scalars c_i, not all zero such that

$$\sum_{i=1}^{p} c_i \mathbf{x}_i = \mathbf{0}.$$

Say $c_k \neq 0$. Then you can solve for $\mathbf{x}_k$ as

$$\mathbf{x}_k = \sum_{j \neq k} (-c_j)/c_k \mathbf{x}_j$$

contrary to assumption. ■

The following is called the exchange theorem.

Theorem 2.4.4 *(Exchange Theorem) Let $\{\mathbf{x}_1, \cdots, \mathbf{x}_r\}$ be a linearly independent set of vectors such that each $\mathbf{x}_i$ is in span$(\mathbf{y}_1, \cdots, \mathbf{y}_s)$. Then $r \leq s$.*

Proof 1: Suppose not. Then $r > s$. By assumption, there exist scalars a_{ji} such that

$$\mathbf{x}_i = \sum_{j=1}^{s} a_{ji} \mathbf{y}_j$$

The matrix whose ji^{th} entry is a_{ji} has more columns than rows. Therefore, by Theorem 2.3.8 there exists a **nonzero** vector $\mathbf{b} \in \mathbb{F}^r$ such that $A\mathbf{b} = \mathbf{0}$. Thus

$$0 = \sum_{i=1}^{r} a_{ji} b_i, \text{ each } j.$$

Then

$$\sum_{i=1}^{r} b_i \mathbf{x}_i = \sum_{i=1}^{r} b_i \sum_{j=1}^{s} a_{ji} \mathbf{y}_j = \sum_{j=1}^{s} \left(\sum_{i=1}^{r} a_{ji} b_i \right) \mathbf{y}_j = \mathbf{0}$$

contradicting the assumption that $\{\mathbf{x}_1, \cdots, \mathbf{x}_r\}$ is linearly independent.

Proof 2: Define $\text{span}\{\mathbf{y}_1, \cdots, \mathbf{y}_s\} \equiv V$, it follows there exist scalars $c_1, \cdots, c_s$ such that

$$\mathbf{x}_1 = \sum_{i=1}^{s} c_i \mathbf{y}_i. \tag{2.23}$$

Not all of these scalars can equal zero because if this were the case, it would follow that $\mathbf{x}_1 = 0$ and so $\{\mathbf{x}_1, \cdots, \mathbf{x}_r\}$ would not be linearly independent. Indeed, if $\mathbf{x}_1 = \mathbf{0}$, $1\mathbf{x}_1 + \sum_{i=2}^{r} 0\mathbf{x}_i = \mathbf{x}_1 = \mathbf{0}$ and so there would exist a nontrivial linear combination of the vectors $\{\mathbf{x}_1, \cdots, \mathbf{x}_r\}$ which equals zero.

Say $c_k \neq 0$. Then solve (2.23) for $\mathbf{y}_k$ and obtain

$$\mathbf{y}_k \in \text{span}\left(\mathbf{x}_1, \overbrace{\mathbf{y}_1, \cdots, \mathbf{y}_{k-1}, \mathbf{y}_{k+1}, \cdots, \mathbf{y}_s}^{\text{s-1 vectors here}} \right).$$

Define $\{\mathbf{z}_1, \cdots, \mathbf{z}_{s-1}\}$ by

$$\{\mathbf{z}_1, \cdots, \mathbf{z}_{s-1}\} \equiv \{\mathbf{y}_1, \cdots, \mathbf{y}_{k-1}, \mathbf{y}_{k+1}, \cdots, \mathbf{y}_s\}$$

Therefore, $\text{span}\{\mathbf{x}_1, \mathbf{z}_1, \cdots, \mathbf{z}_{s-1}\} = V$ because if $\mathbf{v} \in V$, there exist constants $c_1, \cdots, c_s$ such that

$$\mathbf{v} = \sum_{i=1}^{s-1} c_i \mathbf{z}_i + c_s \mathbf{y}_k.$$

Now replace the $\mathbf{y}_k$ in the above with a linear combination of the vectors, $\{\mathbf{x}_1, \mathbf{z}_1, \cdots, \mathbf{z}_{s-1}\}$ to obtain $\mathbf{v} \in \text{span}\{\mathbf{x}_1, \mathbf{z}_1, \cdots, \mathbf{z}_{s-1}\}$. The vector $\mathbf{y}_k$, in the list $\{\mathbf{y}_1, \cdots, \mathbf{y}_s\}$, has now been replaced with the vector $\mathbf{x}_1$ and the resulting modified list of vectors has the same span as the original list of vectors, $\{\mathbf{y}_1, \cdots, \mathbf{y}_s\}$.

Now suppose that $r > s$ and that $\text{span}\{\mathbf{x}_1, \cdots, \mathbf{x}_l, \mathbf{z}_1, \cdots, \mathbf{z}_p\} = V$ where the vectors, $\mathbf{z}_1, \cdots, \mathbf{z}_p$ are each taken from the set, $\{\mathbf{y}_1, \cdots, \mathbf{y}_s\}$ and $l + p = s$. This has now been done for $l = 1$ above. Then since $r > s$, it follows that $l \leq s < r$ and so $l + 1 \leq r$. Therefore, $\mathbf{x}_{l+1}$ is a vector not in the list, $\{\mathbf{x}_1, \cdots, \mathbf{x}_l\}$ and since $\text{span}\{\mathbf{x}_1, \cdots, \mathbf{x}_l, \mathbf{z}_1, \cdots, \mathbf{z}_p\} = V$, there exist scalars c_i and d_j such that

$$\mathbf{x}_{l+1} = \sum_{i=1}^{l} c_i \mathbf{x}_i + \sum_{j=1}^{p} d_j \mathbf{z}_j. \tag{2.24}$$

Now not all the d_j can equal zero because if this were so, it would follow that $\{\mathbf{x}_1, \cdots, \mathbf{x}_r\}$ would be a linearly dependent set because one of the vectors would equal a linear combination of the others. Therefore, (2.24) can be solved for one of the $\mathbf{z}_i$, say $\mathbf{z}_k$, in terms of $\mathbf{x}_{l+1}$ and the other $\mathbf{z}_i$ and just as in the above argument, replace that $\mathbf{z}_i$ with $\mathbf{x}_{l+1}$ to obtain

$$\text{span}\left\{\mathbf{x}_1, \cdots \mathbf{x}_l, \mathbf{x}_{l+1}, \overbrace{\mathbf{z}_1, \cdots \mathbf{z}_{k-1}, \mathbf{z}_{k+1}, \cdots, \mathbf{z}_p}^{\text{p-1 vectors here}}\right\} = V.$$

Continue this way, eventually obtaining

$$\text{span}\{\mathbf{x}_1, \cdots, \mathbf{x}_s\} = V.$$

But then $\mathbf{x}_r \in \text{span}\{\mathbf{x}_1, \cdots, \mathbf{x}_s\}$ contrary to the assumption that $\{\mathbf{x}_1, \cdots, \mathbf{x}_r\}$ is linearly independent. Therefore, $r \leq s$ as claimed.

Proof 3: Suppose $r > s$. Let $\mathbf{z}_k$ denote a vector of $\{\mathbf{y}_1, \cdots, \mathbf{y}_s\}$. Thus there exists j as small as possible such that

$$\text{span}(\mathbf{y}_1, \cdots, \mathbf{y}_s) = \text{span}(\mathbf{x}_1, \cdots, \mathbf{x}_m, \mathbf{z}_1, \cdots, \mathbf{z}_j)$$

where $m + j = s$. It is given that $m = 0$, corresponding to no vectors of $\{\mathbf{x}_1, \cdots, \mathbf{x}_m\}$ and $j = s$, corresponding to all the $\mathbf{y}_k$ results in the above equation holding. If $j > 0$ then $m < s$ and so

$$\mathbf{x}_{m+1} = \sum_{k=1}^{m} a_k \mathbf{x}_k + \sum_{i=1}^{j} b_i \mathbf{z}_i$$

Not all the b_i can equal 0 and so you can solve for one of them in terms of $\mathbf{x}_{m+1}, \mathbf{x}_m, \cdots, \mathbf{x}_1$, and the other $\mathbf{z}_k$. Therefore, there exists

$$\{\mathbf{z}_1, \cdots, \mathbf{z}_{j-1}\} \subseteq \{\mathbf{y}_1, \cdots, \mathbf{y}_s\}$$

such that

$$\text{span}(\mathbf{y}_1, \cdots, \mathbf{y}_s) = \text{span}(\mathbf{x}_1, \cdots, \mathbf{x}_{m+1}, \mathbf{z}_1, \cdots, \mathbf{z}_{j-1})$$

contradicting the choice of j. Hence $j = 0$ and

$$\text{span}(\mathbf{y}_1, \cdots, \mathbf{y}_s) = \text{span}(\mathbf{x}_1, \cdots, \mathbf{x}_s)$$

It follows that

$$\mathbf{x}_{s+1} \in \text{span}(\mathbf{x}_1, \cdots, \mathbf{x}_s)$$

contrary to the assumption the $\mathbf{x}_k$ are linearly independent. Therefore, $r \leq s$ as claimed. ■

Definition 2.4.5 *A finite set of vectors, $\{\mathbf{x}_1, \cdots, \mathbf{x}_r\}$ is a basis for $\mathbb{F}^n$ if* $\text{span}(\mathbf{x}_1, \cdots, \mathbf{x}_r) = \mathbb{F}^n$ *and $\{\mathbf{x}_1, \cdots, \mathbf{x}_r\}$ is linearly independent.*

Corollary 2.4.6 *Let $\{\mathbf{x}_1, \cdots, \mathbf{x}_r\}$ and $\{\mathbf{y}_1, \cdots, \mathbf{y}_s\}$ be two bases*[1] *of $\mathbb{F}^n$. Then $r = s = n$.*

[1] This is the plural form of basis. We could say basiss but it would involve an inordinate amount of hissing as in "The sixth shiek's sixth sheep is sick". This is the reason that bases is used instead of basiss.

Proof: From the exchange theorem, $r \leq s$ and $s \leq r$. Now note the vectors,

$$\mathbf{e}_i = \overbrace{(0, \cdots, 0, 1, 0 \cdots, 0)}^{1 \text{ is in the } i^{th} \text{ slot}}$$

for $i = 1, 2, \cdots, n$ are a basis for $\mathbb{F}^n$. ■

Lemma 2.4.7 *Let* $\{\mathbf{v}_1, \cdots, \mathbf{v}_r\}$ *be a set of vectors. Then* $V \equiv \operatorname{span}(\mathbf{v}_1, \cdots, \mathbf{v}_r)$ *is a subspace.*

Proof: Suppose α, β are two scalars and let $\sum_{k=1}^{r} c_k \mathbf{v}_k$ and $\sum_{k=1}^{r} d_k \mathbf{v}_k$ are two elements of V. What about

$$\alpha \sum_{k=1}^{r} c_k \mathbf{v}_k + \beta \sum_{k=1}^{r} d_k \mathbf{v}_k?$$

Is it also in V?

$$\alpha \sum_{k=1}^{r} c_k \mathbf{v}_k + \beta \sum_{k=1}^{r} d_k \mathbf{v}_k = \sum_{k=1}^{r} (\alpha c_k + \beta d_k) \mathbf{v}_k \in V$$

so the answer is yes. ■

Definition 2.4.8 *A finite set of vectors,* $\{\mathbf{x}_1, \cdots, \mathbf{x}_r\}$ *is a basis for a subspace* V *of* $\mathbb{F}^n$ *if* $\operatorname{span}(\mathbf{x}_1, \cdots, \mathbf{x}_r) = V$ *and* $\{\mathbf{x}_1, \cdots, \mathbf{x}_r\}$ *is linearly independent.*

Corollary 2.4.9 *Let* $\{\mathbf{x}_1, \cdots, \mathbf{x}_r\}$ *and* $\{\mathbf{y}_1, \cdots, \mathbf{y}_s\}$ *be two bases for* V*. Then* $r = s$*.*

Proof: From the exchange theorem, $r \leq s$ and $s \leq r$. ■

Definition 2.4.10 *Let* V *be a subspace of* $\mathbb{F}^n$*. Then* $\dim(V)$ *read as the dimension of* V *is the number of vectors in a basis.*

Of course you should wonder right now whether an arbitrary subspace even has a basis. In fact it does and this is in the next theorem. First, here is an interesting lemma.

Lemma 2.4.11 *Suppose* $\mathbf{v} \notin \operatorname{span}(\mathbf{u}_1, \cdots, \mathbf{u}_k)$ *and* $\{\mathbf{u}_1, \cdots, \mathbf{u}_k\}$ *is linearly independent. Then* $\{\mathbf{u}_1, \cdots, \mathbf{u}_k, \mathbf{v}\}$ *is also linearly independent.*

Proof: Suppose $\sum_{i=1}^{k} c_i \mathbf{u}_i + d\mathbf{v} = \mathbf{0}$. It is required to verify that each $c_i = 0$ and that $d = 0$. But if $d \neq 0$, then you can solve for $\mathbf{v}$ as a linear combination of the vectors, $\{\mathbf{u}_1, \cdots, \mathbf{u}_k\}$,

$$\mathbf{v} = -\sum_{i=1}^{k} \left(\frac{c_i}{d}\right) \mathbf{u}_i$$

contrary to assumption. Therefore, $d = 0$. But then $\sum_{i=1}^{k} c_i \mathbf{u}_i = 0$ and the linear independence of $\{\mathbf{u}_1, \cdots, \mathbf{u}_k\}$ implies each $c_i = 0$ also. ■

Theorem 2.4.12 *Let* V *be a nonzero subspace of* $\mathbb{F}^n$*. Then* V *has a basis.*

Proof: Let $\mathbf{v}_1 \in V$ where $\mathbf{v}_1 \neq 0$. If $\operatorname{span}\{\mathbf{v}_1\} = V$, stop. $\{\mathbf{v}_1\}$ is a basis for V. Otherwise, there exists $\mathbf{v}_2 \in V$ which is not in $\operatorname{span}\{\mathbf{v}_1\}$. By Lemma 2.4.11 $\{\mathbf{v}_1, \mathbf{v}_2\}$ is a linearly independent set of vectors. If $\operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2\} = V$ stop, $\{\mathbf{v}_1, \mathbf{v}_2\}$ is a basis for V. If $\operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2\} \neq V$, then there exists $\mathbf{v}_3 \notin \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2\}$ and $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is a larger linearly independent set of vectors. Continuing this way, the process must stop before $n + 1$ steps because if not, it would be possible to obtain $n + 1$ linearly independent vectors contrary to the exchange theorem. ■

In words the following corollary states that any linearly independent set of vectors can be enlarged to form a basis.

Corollary 2.4.13 *Let* V *be a subspace of* $\mathbb{F}^n$ *and let* $\{\mathbf{v}_1, \cdots, \mathbf{v}_r\}$ *be a linearly independent set of vectors in* V*. Then either it is a basis for* V *or there exist vectors,* $\mathbf{v}_{r+1}, \cdots, \mathbf{v}_s$ *such that* $\{\mathbf{v}_1, \cdots, \mathbf{v}_r, \mathbf{v}_{r+1}, \cdots, \mathbf{v}_s\}$ *is a basis for* V*.*

Proof: This follows immediately from the proof of Theorem 2.4.12. You do exactly the same argument except you start with $\{\mathbf{v}_1, \cdots, \mathbf{v}_r\}$ rather than $\{\mathbf{v}_1\}$. ■

It is also true that any spanning set of vectors can be restricted to obtain a basis.

Theorem 2.4.14 *Let* V *be a subspace of* $\mathbb{F}^n$ *and suppose* $\operatorname{span}(\mathbf{u}_1 \cdots, \mathbf{u}_p) = V$ *where the* $\mathbf{u}_i$ *are nonzero vectors. Then there exist vectors* $\{\mathbf{v}_1 \cdots, \mathbf{v}_r\}$ *such that* $\{\mathbf{v}_1 \cdots, \mathbf{v}_r\} \subseteq \{\mathbf{u}_1 \cdots, \mathbf{u}_p\}$ *and* $\{\mathbf{v}_1 \cdots, \mathbf{v}_r\}$ *is a basis for* V*.*

Proof: Let r be the smallest positive integer with the property that for some set $\{\mathbf{v}_1 \cdots, \mathbf{v}_r\} \subseteq \{\mathbf{u}_1 \cdots, \mathbf{u}_p\}$,

$$\operatorname{span}(\mathbf{v}_1 \cdots, \mathbf{v}_r) = V.$$

Then $r \leq p$ and it must be the case that $\{\mathbf{v}_1 \cdots, \mathbf{v}_r\}$ is linearly independent because if it were not so, one of the vectors, say $\mathbf{v}_k$ would be a linear combination of the others. But then you could delete this vector from $\{\mathbf{v}_1 \cdots, \mathbf{v}_r\}$ and the resulting list of $r - 1$ vectors would still span V contrary to the definition of r. ■

2.5 An Application To Matrices

The following is a theorem of major significance.

Theorem 2.5.1 *Suppose A is an $n \times n$ matrix. Then A is one to one (injective) if and only if A is onto (surjective). Also, if B is an $n \times n$ matrix and $AB = I$, then it follows $BA = I$.*

Proof: First suppose A is one to one. Consider the vectors, $\{A\mathbf{e}_1, \cdots, A\mathbf{e}_n\}$ where $\mathbf{e}_k$ is the column vector which is all zeros except for a 1 in the k^{th} position. This set of vectors is linearly independent because if

$$\sum_{k=1}^{n} c_k A\mathbf{e}_k = \mathbf{0},$$

then since A is linear,

$$A\left(\sum_{k=1}^{n} c_k \mathbf{e}_k\right) = \mathbf{0}$$

and since A is one to one, it follows

$$\sum_{k=1}^{n} c_k \mathbf{e}_k = \mathbf{0}$$

which implies each $c_k = 0$ because the $\mathbf{e}_k$ are clearly linearly independent.

Therefore, $\{A\mathbf{e}_1, \cdots, A\mathbf{e}_n\}$ must be a basis for $\mathbb{F}^n$ because if not there would exist a vector, $\mathbf{y} \notin \operatorname{span}(A\mathbf{e}_1, \cdots, A\mathbf{e}_n)$ and then by Lemma 2.4.11, $\{A\mathbf{e}_1, \cdots, A\mathbf{e}_n, \mathbf{y}\}$ would be an independent set of vectors having $n+1$ vectors in it, contrary to the exchange theorem. It follows that for $\mathbf{y} \in \mathbb{F}^n$ there exist constants, c_i such that

$$\mathbf{y} = \sum_{k=1}^{n} c_k A\mathbf{e}_k = A\left(\sum_{k=1}^{n} c_k \mathbf{e}_k\right)$$

showing that, since $\mathbf{y}$ was arbitrary, A is onto.

Next suppose A is onto. This means the span of the columns of A equals $\mathbb{F}^n$. If these columns are not linearly independent, then by Lemma 2.4.3 on Page 74, one of the columns is a linear combination of the others and so the span of the columns of A equals the span of the $n-1$ other columns. This violates the exchange theorem because $\{\mathbf{e}_1, \cdots, \mathbf{e}_n\}$ would be a linearly independent set of vectors contained in the span of only $n-1$ vectors. Therefore, the columns of A must be independent and this is equivalent to saying that $A\mathbf{x} = \mathbf{0}$ if and only if $\mathbf{x} = \mathbf{0}$. This implies A is one to one because if $A\mathbf{x} = A\mathbf{y}$, then $A(\mathbf{x} - \mathbf{y}) = \mathbf{0}$ and so $\mathbf{x} - \mathbf{y} = \mathbf{0}$.

Now suppose $AB = I$. Why is $BA = I$? Since $AB = I$ it follows B is one to one since otherwise, there would exist, $\mathbf{x} \neq \mathbf{0}$ such that $B\mathbf{x} = \mathbf{0}$ and then $AB\mathbf{x} = A\mathbf{0} = \mathbf{0} \neq I\mathbf{x}$. Therefore, from what was just shown, B is also onto. In addition to this, A must be one to one because if $A\mathbf{y} = \mathbf{0}$, then $\mathbf{y} = B\mathbf{x}$ for some $\mathbf{x}$ and then $\mathbf{x} = AB\mathbf{x} = A\mathbf{y} = \mathbf{0}$ showing $\mathbf{y} = \mathbf{0}$. Now from what is given to be so, it follows $(AB)A = A$ and so using the associative law for matrix multiplication,

$$A(BA) - A = A(BA - I) = 0.$$

But this means $(BA - I)\mathbf{x} = \mathbf{0}$ for all $\mathbf{x}$ since otherwise, A would not be one to one. Hence $BA = I$ as claimed. ■

This theorem shows that if an $n \times n$ matrix B acts like an inverse when multiplied on one side of A, it follows that $B = A^{-1}$and it will act like an inverse on both sides of A.

The conclusion of this theorem pertains to square matrices only. For example, let

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}, \ B = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & -1 \end{pmatrix} \tag{2.25}$$

Then

$$BA = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

but

$$AB = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & -1 \\ 1 & 0 & 0 \end{pmatrix}.$$

2.6 Matrices And Calculus

The study of moving coordinate systems gives a non trivial example of the usefulness of the ideas involving linear transformations and matrices. To begin with, here is the concept of the product rule extended to matrix multiplication.

Definition 2.6.1 *Let $A(t)$ be an $m \times n$ matrix. Say $A(t) = (A_{ij}(t))$. Suppose also that $A_{ij}(t)$ is a differentiable function for all i, j. Then define $A'(t) \equiv \left(A'_{ij}(t)\right)$. That is, $A'(t)$ is the matrix which consists of replacing each entry by its derivative. Such an $m \times n$ matrix in which the entries are differentiable functions is called a differentiable matrix.*

The next lemma is just a version of the product rule.

Lemma 2.6.2 *Let $A(t)$ be an $m \times n$ matrix and let $B(t)$ be an $n \times p$ matrix with the property that all the entries of these matrices are differentiable functions. Then*

$$(A(t) B(t))' = A'(t) B(t) + A(t) B'(t).$$

Proof: This is like the usual proof.

$$\frac{1}{h}(A(t+h) B(t+h) - A(t) B(t)) =$$

$$\frac{1}{h}(A(t+h) B(t+h) - A(t+h) B(t)) + \frac{1}{h}(A(t+h) B(t) - A(t) B(t))$$

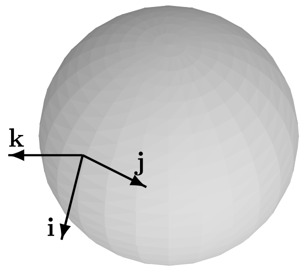
$$= A(t+h) \frac{B(t+h) - B(t)}{h} + \frac{A(t+h) - A(t)}{h} B(t)$$

and now, using the fact that the entries of the matrices are all differentiable, one can pass to a limit in both sides as $h \to 0$ and conclude that

$$(A(t) B(t))' = A'(t) B(t) + A(t) B'(t) \blacksquare$$

2.6.1 The Coriolis Acceleration

Imagine a point on the surface of the earth. Now consider unit vectors, one pointing South, one pointing East and one pointing directly away from the center of the earth.



Denote the first as $\mathbf{i}$, the second as $\mathbf{j}$, and the third as $\mathbf{k}$. If you are standing on the earth you will consider these vectors as fixed, but of course they are not. As the earth turns, they change direction and so each is in reality a function of t. Nevertheless, it is with respect to these apparently fixed vectors that you wish to understand acceleration, velocities, and displacements.

In general, let $\mathbf{i}^*, \mathbf{j}^*, \mathbf{k}^*$ be the usual fixed vectors in space and let $\mathbf{i}(t), \mathbf{j}(t), \mathbf{k}(t)$ be an orthonormal basis of vectors for each t, like the vectors described in the first paragraph. It is assumed these vectors are C^1 functions of t. Letting the positive x axis extend in the direction of $\mathbf{i}(t)$, the positive y axis extend in the direction of $\mathbf{j}(t)$, and the positive z axis extend in the direction of $\mathbf{k}(t)$, yields a moving coordinate system. Now let $\mathbf{u}$ be a vector and let t_0 be some reference time. For example you could let $t_0 = 0$. Then define the components of $\mathbf{u}$ with respect to these vectors, $\mathbf{i}, \mathbf{j}, \mathbf{k}$ at time t_0 as

$$\mathbf{u} \equiv u^1 \mathbf{i}(t_0) + u^2 \mathbf{j}(t_0) + u^3 \mathbf{k}(t_0).$$

Let $\mathbf{u}(t)$ be defined as the vector which has the same components with respect to $\mathbf{i}, \mathbf{j}, \mathbf{k}$ but at time t. Thus

$$\mathbf{u}(t) \equiv u^1 \mathbf{i}(t) + u^2 \mathbf{j}(t) + u^3 \mathbf{k}(t).$$

and the vector has changed although the components have not.

This is exactly the situation in the case of the apparently fixed basis vectors on the earth if $\mathbf{u}$ is a position vector from the given spot on the earth's surface to a point regarded as fixed with the earth due to its keeping the same coordinates relative to the coordinate axes which are fixed with the earth. Now define a linear transformation $Q(t)$ mapping $\mathbb{R}^3$ to $\mathbb{R}^3$ by

$$Q(t)\,\mathbf{u} \equiv u^1 \mathbf{i}(t) + u^2 \mathbf{j}(t) + u^3 \mathbf{k}(t)$$

where

$$\mathbf{u} \equiv u^1 \mathbf{i}(t_0) + u^2 \mathbf{j}(t_0) + u^3 \mathbf{k}(t_0)$$

Thus letting $\mathbf{v}$ be a vector defined in the same manner as $\mathbf{u}$ and α, β, scalars,

$$Q(t)(\alpha\mathbf{u} + \beta\mathbf{v}) \equiv \left(\alpha u^1 + \beta v^1\right)\mathbf{i}(t) + \left(\alpha u^2 + \beta v^2\right)\mathbf{j}(t) + \left(\alpha u^3 + \beta v^3\right)\mathbf{k}(t)$$

$$\begin{aligned}
&= \left(\alpha u^1 \mathbf{i}(t) + \alpha u^2 \mathbf{j}(t) + \alpha u^3 \mathbf{k}(t)\right) + \left(\beta v^1 \mathbf{i}(t) + \beta v^2 \mathbf{j}(t) + \beta v^3 \mathbf{k}(t)\right) \\
&= \alpha\left(u^1 \mathbf{i}(t) + u^2 \mathbf{j}(t) + u^3 \mathbf{k}(t)\right) + \beta\left(v^1 \mathbf{i}(t) + v^2 \mathbf{j}(t) + v^3 \mathbf{k}(t)\right) \\
&\equiv \alpha Q(t)\,\mathbf{u} + \beta Q(t)\,\mathbf{v}
\end{aligned}$$

showing that $Q(t)$ is a linear transformation. Also, $Q(t)$ preserves all distances because, since the vectors, $\mathbf{i}(t), \mathbf{j}(t), \mathbf{k}(t)$ form an orthonormal set,

$$|Q(t)\mathbf{u}| = \left(\sum_{i=1}^{3} \left(u^i\right)^2\right)^{1/2} = |\mathbf{u}|.$$

Lemma 2.6.3 *Suppose $Q(t)$ is a real, differentiable $n\times n$ matrix which preserves distances. Then $Q(t)Q(t)^T = Q(t)^T Q(t) = I$. Also, if $\mathbf{u}(t) \equiv Q(t)\mathbf{u}$, then there exists a vector, $\mathbf{\Omega}(t)$ such that*

$$\mathbf{u}'(t) = \mathbf{\Omega}(t) \times \mathbf{u}(t).$$

The symbol $\times$ refers to the cross product.

Proof: Recall that $(\mathbf{z}\cdot\mathbf{w}) = \frac{1}{4}\left(|\mathbf{z}+\mathbf{w}|^2 - |\mathbf{z}-\mathbf{w}|^2\right)$. Therefore,

$$\begin{aligned}(Q(t)\mathbf{u}\cdot Q(t)\mathbf{w}) &= \frac{1}{4}\left(|Q(t)(\mathbf{u}+\mathbf{w})|^2 - |Q(t)(\mathbf{u}-\mathbf{w})|^2\right) \\ &= \frac{1}{4}\left(|\mathbf{u}+\mathbf{w}|^2 - |\mathbf{u}-\mathbf{w}|^2\right) \\ &= (\mathbf{u}\cdot\mathbf{w}).\end{aligned}$$

This implies

$$\left(Q(t)^T Q(t)\mathbf{u}\cdot\mathbf{w}\right) = (\mathbf{u}\cdot\mathbf{w})$$

for all $\mathbf{u},\mathbf{w}$. Therefore, $Q(t)^T Q(t)\mathbf{u} = \mathbf{u}$ and so $Q(t)^T Q(t) = Q(t)Q(t)^T = I$. This proves the first part of the lemma.

It follows from the product rule, Lemma 2.6.2 that

$$Q'(t)Q(t)^T + Q(t)Q'(t)^T = 0$$

and so

$$Q'(t)Q(t)^T = -\left(Q'(t)Q(t)^T\right)^T. \tag{2.26}$$

From the definition, $Q(t)\mathbf{u} = \mathbf{u}(t)$,

$$\mathbf{u}'(t) = Q'(t)\mathbf{u} = Q'(t)\overbrace{Q(t)^T\mathbf{u}(t)}^{=\mathbf{u}}.$$

Then writing the matrix of $Q'(t)Q(t)^T$ with respect to fixed in space orthonormal basis vectors, $\mathbf{i}^*, \mathbf{j}^*, \mathbf{k}^*$, where these are the usual basis vectors for $\mathbb{R}^3$, it follows from 2.26 that the matrix of $Q'(t)Q(t)^T$ is of the form

$$\begin{pmatrix} 0 & -\omega_3(t) & \omega_2(t) \\ \omega_3(t) & 0 & -\omega_1(t) \\ -\omega_2(t) & \omega_1(t) & 0 \end{pmatrix}$$

for some time dependent scalars ω_i. Therefore,

$$\begin{pmatrix} u^1 \\ u^2 \\ u^3 \end{pmatrix}'(t) = \begin{pmatrix} 0 & -\omega_3(t) & \omega_2(t) \\ \omega_3(t) & 0 & -\omega_1(t) \\ -\omega_2(t) & \omega_1(t) & 0 \end{pmatrix}\begin{pmatrix} u^1 \\ u^2 \\ u^3 \end{pmatrix}(t)$$

where the u^i are the components of the vector $\mathbf{u}(t)$ in terms of the fixed vectors $\mathbf{i}^*, \mathbf{j}^*, \mathbf{k}^*$. Therefore,

$$\mathbf{u}'(t) = \mathbf{\Omega}(t) \times \mathbf{u}(t) = Q'(t) Q(t)^T \mathbf{u}(t) \tag{2.27}$$

where

$$\mathbf{\Omega}(t) = \omega_1(t)\,\mathbf{i}^* + \omega_2(t)\,\mathbf{j}^* + \omega_3(t)\,\mathbf{k}^*.$$

because

$$\mathbf{\Omega}(t) \times \mathbf{u}(t) \equiv \begin{vmatrix} \mathbf{i}^* & \mathbf{j}^* & \mathbf{k}^* \\ w_1 & w_2 & w_3 \\ u^1 & u^2 & u^3 \end{vmatrix} \equiv$$

$$\mathbf{i}^*\left(w_2 u^3 - w_3 u^2\right) + \mathbf{j}^*\left(w_3 u^1 - w_1^3\right) + \mathbf{k}^*\left(w_1 u^2 - w_2 u^1\right).$$

This proves the lemma and yields the existence part of the following theorem. ■

Theorem 2.6.4 *Let $\mathbf{i}(t), \mathbf{j}(t), \mathbf{k}(t)$ be as described. Then there exists a unique vector $\mathbf{\Omega}(t)$ such that if $\mathbf{u}(t)$ is a vector whose components are constant with respect to $\mathbf{i}(t), \mathbf{j}(t), \mathbf{k}(t)$, then*

$$\mathbf{u}'(t) = \mathbf{\Omega}(t) \times \mathbf{u}(t).$$

Proof: It only remains to prove uniqueness. Suppose $\mathbf{\Omega}_1$ also works. Then $\mathbf{u}(t) = Q(t)\,\mathbf{u}$ and so $\mathbf{u}'(t) = Q'(t)\,\mathbf{u}$ and

$$Q'(t)\,\mathbf{u} = \mathbf{\Omega} \times Q(t)\,\mathbf{u} = \mathbf{\Omega}_1 \times Q(t)\,\mathbf{u}$$

for all $\mathbf{u}$. Therefore,

$$(\mathbf{\Omega} - \mathbf{\Omega}_1) \times Q(t)\,\mathbf{u} = \mathbf{0}$$

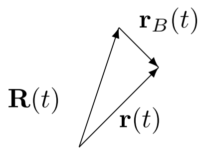
for all $\mathbf{u}$ and since $Q(t)$ is one to one and onto, this implies $(\mathbf{\Omega} - \mathbf{\Omega}_1) \times \mathbf{w} = \mathbf{0}$ for all $\mathbf{w}$ and thus $\mathbf{\Omega} - \mathbf{\Omega}_1 = \mathbf{0}$. ■

Now let $\mathbf{R}(t)$ be a position vector and let

$$\mathbf{r}(t) = \mathbf{R}(t) + \mathbf{r}_B(t)$$

where

$$\mathbf{r}_B(t) \equiv x(t)\,\mathbf{i}(t) + y(t)\,\mathbf{j}(t) + z(t)\,\mathbf{k}(t).$$



In the example of the earth, $\mathbf{R}(t)$ is the position vector of a point $\mathbf{p}(t)$ on the earth's surface and $\mathbf{r}_B(t)$ is the position vector of another point from $\mathbf{p}(t)$, thus regarding $\mathbf{p}(t)$ as the origin. $\mathbf{r}_B(t)$ is the position vector of a point as perceived by the observer on the earth with respect to the vectors he thinks of as fixed. Similarly, $\mathbf{v}_B(t)$ and $\mathbf{a}_B(t)$ will be the velocity and acceleration relative to $\mathbf{i}(t), \mathbf{j}(t), \mathbf{k}(t)$, and so $\mathbf{v}_B = x'\mathbf{i} + y'\mathbf{j} + z'\mathbf{k}$ and $\mathbf{a}_B = x''\mathbf{i} + y''\mathbf{j} + z''\mathbf{k}$. Then

$$\mathbf{v} \equiv \mathbf{r}' = \mathbf{R}' + x'\mathbf{i} + y'\mathbf{j} + z'\mathbf{k} + x\mathbf{i}' + y\mathbf{j}' + z\mathbf{k}'.$$

By , 2.27, if $\mathbf{e} \in \{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$, $\mathbf{e}' = \mathbf{\Omega} \times \mathbf{e}$ because the components of these vectors with respect to $\mathbf{i}, \mathbf{j}, \mathbf{k}$ are constant. Therefore,

$$\begin{aligned} x\mathbf{i}' + y\mathbf{j}' + z\mathbf{k}' &= x\mathbf{\Omega} \times \mathbf{i} + y\mathbf{\Omega} \times \mathbf{j} + z\mathbf{\Omega} \times \mathbf{k} \\ &= \mathbf{\Omega} \times (x\mathbf{i} + y\mathbf{j} + z\mathbf{k}) \end{aligned}$$

and consequently,

$$\mathbf{v} = \mathbf{R}' + x'\mathbf{i} + y'\mathbf{j} + z'\mathbf{k} + \mathbf{\Omega} \times \mathbf{r}_B = \mathbf{R}' + x'\mathbf{i} + y'\mathbf{j} + z'\mathbf{k} + \mathbf{\Omega} \times (x\mathbf{i} + y\mathbf{j} + z\mathbf{k}).$$

Now consider the acceleration. Quantities which are relative to the moving coordinate system and quantities which are relative to a fixed coordinate system are distinguished by using the subscript B on those relative to the moving coordinate system.

$$\begin{aligned}\mathbf{a} = \mathbf{v}' = \mathbf{R}'' + x''\mathbf{i} + y''\mathbf{j} + z''\mathbf{k} + \overbrace{x'\mathbf{i}' + y'\mathbf{j}' + z'\mathbf{k}'}^{\boldsymbol{\Omega}\times\mathbf{v}_B} + \ \boldsymbol{\Omega}' \times \mathbf{r}_B \\ +\boldsymbol{\Omega}\times\left(\overbrace{x'\mathbf{i} + y'\mathbf{j} + z'\mathbf{k}}^{\mathbf{v}_B} + \overbrace{x\mathbf{i}' + y\mathbf{j}' + z\mathbf{k}'}^{\boldsymbol{\Omega}\times\mathbf{r}_B(t)}\right) \\ = \mathbf{R}'' + \mathbf{a}_B + \boldsymbol{\Omega}' \times \mathbf{r}_B + 2\boldsymbol{\Omega} \times \mathbf{v}_B + \boldsymbol{\Omega}\times(\boldsymbol{\Omega} \times \mathbf{r}_B).\end{aligned}$$

The acceleration $\mathbf{a}_B$ is that perceived by an observer who is moving with the moving coordinate system and for whom the moving coordinate system is fixed. The term $\boldsymbol{\Omega}\times(\boldsymbol{\Omega} \times \mathbf{r}_B)$ is called the centripetal acceleration. Solving for $\mathbf{a}_B$,

$$\mathbf{a}_B = \mathbf{a} - \mathbf{R}'' - \boldsymbol{\Omega}' \times \mathbf{r}_B - 2\boldsymbol{\Omega} \times \mathbf{v}_B - \boldsymbol{\Omega}\times(\boldsymbol{\Omega} \times \mathbf{r}_B). \tag{2.28}$$

Here the term $-(\boldsymbol{\Omega}\times(\boldsymbol{\Omega} \times \mathbf{r}_B))$ is called the centrifugal acceleration, it being an acceleration felt by the observer relative to the moving coordinate system which he regards as fixed, and the term $-2\boldsymbol{\Omega} \times \mathbf{v}_B$ is called the Coriolis acceleration, an acceleration experienced by the observer as he moves relative to the moving coordinate system. The mass multiplied by the Coriolis acceleration defines the Coriolis force.

There is a ride found in some amusement parks in which the victims stand next to a circular wall covered with a carpet or some rough material. Then the whole circular room begins to revolve faster and faster. At some point, the bottom drops out and the victims are held in place by friction. The force they feel is called centrifugal force and it causes centrifugal acceleration. It is not necessary to move relative to coordinates fixed with the revolving wall in order to feel this force and it is pretty predictable. However, if the nauseated victim moves relative to the rotating wall, he will feel the effects of the Coriolis force and this force is really strange. The difference between these forces is that the Coriolis force is caused by movement relative to the moving coordinate system and the centrifugal force is not.

2.6.2 The Coriolis Acceleration On The Rotating Earth

Now consider the earth. Let $\mathbf{i}^*, \mathbf{j}^*, \mathbf{k}^*$, be the usual basis vectors fixed in space with $\mathbf{k}^*$ pointing in the direction of the north pole from the center of the earth and let $\mathbf{i}, \mathbf{j}, \mathbf{k}$ be the unit vectors described earlier with $\mathbf{i}$ pointing South, $\mathbf{j}$ pointing East, and $\mathbf{k}$ pointing away from the center of the earth at some point of the rotating earth's surface $\mathbf{p}$. Letting $\mathbf{R}(t)$ be the position vector of the point $\mathbf{p}$, from the center of the earth, observe the coordinates of $\mathbf{R}(t)$ are constant with respect to $\mathbf{i}(t), \mathbf{j}(t), \mathbf{k}(t)$. Also, since the earth rotates from West to East and the speed of a point on the surface of the earth relative to an observer fixed in space is $\omega|\mathbf{R}|\sin\phi$ where ω is the angular speed of the earth about an axis through the poles and ϕ is the polar angle measured from the positive z axis down as in spherical coordinates. It follows from the geometric definition of the cross product that

$$\mathbf{R}' = \omega\mathbf{k}^* \times \mathbf{R}$$

Therefore, the vector of Theorem 2.6.4 is $\boldsymbol{\Omega} = \omega\mathbf{k}^*$ and so

$$\mathbf{R}'' = \overbrace{\boldsymbol{\Omega}' \times \mathbf{R}}^{=\mathbf{0}} + \ \boldsymbol{\Omega} \times \mathbf{R}' = \boldsymbol{\Omega}\times(\boldsymbol{\Omega} \times \mathbf{R})$$

since $\mathbf{\Omega}$ does not depend on t. Formula 2.28 implies

$$\mathbf{a}_B = \mathbf{a} - \mathbf{\Omega}\times(\mathbf{\Omega}\times\mathbf{R}) - 2\mathbf{\Omega}\times\mathbf{v}_B - \mathbf{\Omega}\times(\mathbf{\Omega}\times\mathbf{r}_B). \tag{2.29}$$

In this formula, you can totally ignore the term $\mathbf{\Omega}\times(\mathbf{\Omega}\times\mathbf{r}_B)$ because it is so small whenever you are considering motion near some point on the earth's surface. To see this, note $\omega\overbrace{(24)(3600)}^{\text{seconds in a day}} = 2\pi$, and so $\omega = 7.2722\times10^{-5}$ in radians per second. If you are using seconds to measure time and feet to measure distance, this term is therefore, no larger than

$$\left(7.2722\times10^{-5}\right)^2 |\mathbf{r}_B|.$$

Clearly this is not worth considering in the presence of the acceleration due to gravity which is approximately 32 feet per second squared near the surface of the earth.

If the acceleration $\mathbf{a}$ is due to gravity, then

$$\mathbf{a}_B = \mathbf{a} - \mathbf{\Omega}\times(\mathbf{\Omega}\times\mathbf{R}) - 2\mathbf{\Omega}\times\mathbf{v}_B =$$

$$\overbrace{-\frac{GM(\mathbf{R}+\mathbf{r}_B)}{|\mathbf{R}+\mathbf{r}_B|^3} - \mathbf{\Omega}\times(\mathbf{\Omega}\times\mathbf{R})}^{\equiv\mathbf{g}} - 2\mathbf{\Omega}\times\mathbf{v}_B \equiv \mathbf{g} - 2\mathbf{\Omega}\times\mathbf{v_B}.$$

Note that

$$\mathbf{\Omega}\times(\mathbf{\Omega}\times\mathbf{R}) = (\mathbf{\Omega}\cdot\mathbf{R})\,\mathbf{\Omega} - |\mathbf{\Omega}|^2\,\mathbf{R}$$

and so $\mathbf{g}$, the acceleration relative to the moving coordinate system on the earth is not directed exactly toward the center of the earth except at the poles and at the equator, although the components of acceleration which are in other directions are very small when compared with the acceleration due to the force of gravity and are often neglected. Therefore, if the only force acting on an object is due to gravity, the following formula describes the acceleration relative to a coordinate system moving with the earth's surface.

$$\mathbf{a}_B = \mathbf{g} - 2(\mathbf{\Omega}\times\mathbf{v}_B)$$

While the vector $\mathbf{\Omega}$ is quite small, if the relative velocity, $\mathbf{v}_B$ is large, the Coriolis acceleration could be significant. This is described in terms of the vectors $\mathbf{i}(t), \mathbf{j}(t), \mathbf{k}(t)$ next.

Letting (ρ,θ,ϕ) be the usual spherical coordinates of the point $\mathbf{p}(t)$ on the surface taken with respect to $\mathbf{i}^*, \mathbf{j}^*, \mathbf{k}^*$ the usual way with ϕ the polar angle, it follows the $\mathbf{i}^*, \mathbf{j}^*, \mathbf{k}^*$ coordinates of this point are

$$\begin{pmatrix} \rho\sin(\phi)\cos(\theta) \\ \rho\sin(\phi)\sin(\theta) \\ \rho\cos(\phi) \end{pmatrix}.$$

It follows,

$$\mathbf{i} = \cos(\phi)\cos(\theta)\,\mathbf{i}^* + \cos(\phi)\sin(\theta)\,\mathbf{j}^* - \sin(\phi)\,\mathbf{k}^*$$
$$\mathbf{j} = -\sin(\theta)\,\mathbf{i}^* + \cos(\theta)\,\mathbf{j}^* + 0\mathbf{k}^*$$

and

$$\mathbf{k} = \sin(\phi)\cos(\theta)\,\mathbf{i}^* + \sin(\phi)\sin(\theta)\,\mathbf{j}^* + \cos(\phi)\,\mathbf{k}^*.$$

It is necessary to obtain $\mathbf{k}^*$ in terms of the vectors, $\mathbf{i},\mathbf{j},\mathbf{k}$. Thus the following equation needs to be solved for a,b,c to find $\mathbf{k}^* = a\mathbf{i}+b\mathbf{j}+c\mathbf{k}$

$$\overbrace{\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}}^{\mathbf{k}^*} = \begin{pmatrix} \cos(\phi)\cos(\theta) & -\sin(\theta) & \sin(\phi)\cos(\theta) \\ \cos(\phi)\sin(\theta) & \cos(\theta) & \sin(\phi)\sin(\theta) \\ -\sin(\phi) & 0 & \cos(\phi) \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} \tag{2.30}$$

The first column is $\mathbf{i}$, the second is $\mathbf{j}$ and the third is $\mathbf{k}$ in the above matrix. The solution is $a = -\sin(\phi)$, $b = 0$, and $c = \cos(\phi)$.

Now the Coriolis acceleration on the earth equals

$$2(\mathbf{\Omega} \times \mathbf{v}_B) = 2\omega \left(\overbrace{-\sin(\phi)\,\mathbf{i}+0\mathbf{j}+\cos(\phi)\,\mathbf{k}}^{\mathbf{k}^*} \right) \times (x'\mathbf{i}+y'\mathbf{j}+z'\mathbf{k}).$$

This equals

$$2\omega\left[(-y'\cos\phi)\,\mathbf{i}+(x'\cos\phi+z'\sin\phi)\,\mathbf{j}-(y'\sin\phi)\,\mathbf{k}\right]. \tag{2.31}$$

Remember ϕ is fixed and pertains to the fixed point, $\mathbf{p}(t)$ on the earth's surface. Therefore, if the acceleration $\mathbf{a}$ is due to gravity,

$$\mathbf{a}_B = \mathbf{g} - 2\omega\left[(-y'\cos\phi)\,\mathbf{i}+(x'\cos\phi+z'\sin\phi)\,\mathbf{j}-(y'\sin\phi)\,\mathbf{k}\right]$$

where $\mathbf{g} = -\frac{GM(\mathbf{R}+\mathbf{r}_B)}{|\mathbf{R}+\mathbf{r}_B|^3} - \mathbf{\Omega}\times(\mathbf{\Omega}\times\mathbf{R})$ as explained above. The term $\mathbf{\Omega}\times(\mathbf{\Omega}\times\mathbf{R})$ is pretty small and so it will be neglected. However, the Coriolis force will not be neglected.

Example 2.6.5 *Suppose a rock is dropped from a tall building. Where will it strike?*

Assume $\mathbf{a} = -g\mathbf{k}$ and the $\mathbf{j}$ component of $\mathbf{a}_B$ is approximately

$$-2\omega\left(x'\cos\phi + z'\sin\phi\right).$$

The dominant term in this expression is clearly the second one because x' will be small. Also, the $\mathbf{i}$ and $\mathbf{k}$ contributions will be very small. Therefore, the following equation is descriptive of the situation.

$$\mathbf{a}_B = -g\mathbf{k}-2z'\omega\sin\phi\mathbf{j}.$$

$z' = -gt$ approximately. Therefore, considering the $\mathbf{j}$ component, this is

$$2gt\omega\sin\phi.$$

Two integrations give $\left(\omega g t^3/3\right)\sin\phi$ for the $\mathbf{j}$ component of the relative displacement at time t.

This shows the rock does not fall directly towards the center of the earth as expected but slightly to the east.

Example 2.6.6 *In 1851 Foucault set a pendulum vibrating and observed the earth rotate out from under it. It was a very long pendulum with a heavy weight at the end so that it would vibrate for a long time without stopping[2]. This is what allowed him to observe the earth rotate out from under it. Clearly such a pendulum will take 24 hours for the plane of vibration to appear to make one complete revolution at the north pole. It is also reasonable to expect that no such observed rotation would take place on the equator. Is it possible to predict what will take place at various latitudes?*

Using 2.31, in 2.29,

$$\mathbf{a}_B = \mathbf{a} - \mathbf{\Omega}\times(\mathbf{\Omega}\times\mathbf{R})$$
$$-2\omega\left[(-y'\cos\phi)\,\mathbf{i}+(x'\cos\phi+z'\sin\phi)\,\mathbf{j}-(y'\sin\phi)\,\mathbf{k}\right].$$

Neglecting the small term, $\mathbf{\Omega}\times(\mathbf{\Omega}\times\mathbf{R})$, this becomes

$$= -g\mathbf{k}+\mathbf{T}/m-2\omega\left[(-y'\cos\phi)\,\mathbf{i}+(x'\cos\phi+z'\sin\phi)\,\mathbf{j}-(y'\sin\phi)\,\mathbf{k}\right]$$

where $\mathbf{T}$, the tension in the string of the pendulum, is directed towards the point at which the pendulum is supported, and m is the mass of the pendulum bob. The pendulum can be thought of as the position vector from $(0,0,l)$ to the surface of the sphere $x^2+y^2+(z-l)^2 = l^2$. Therefore,

$$\mathbf{T} = -T\frac{x}{l}\mathbf{i}-T\frac{y}{l}\mathbf{j}+T\frac{l-z}{l}\mathbf{k}$$

and consequently, the differential equations of relative motion are

$$x'' = -T\frac{x}{ml}+2\omega y'\cos\phi$$

$$y'' = -T\frac{y}{ml}-2\omega\left(x'\cos\phi+z'\sin\phi\right)$$

and

$$z'' = T\frac{l-z}{ml}-g+2\omega y'\sin\phi.$$

If the vibrations of the pendulum are small so that for practical purposes, $z''=z=0$, the last equation may be solved for T to get

$$gm-2\omega y'\sin(\phi)\,m = T.$$

Therefore, the first two equations become

$$x'' = -\left(gm-2\omega m y'\sin\phi\right)\frac{x}{ml}+2\omega y'\cos\phi$$

and

$$y'' = -\left(gm-2\omega m y'\sin\phi\right)\frac{y}{ml}-2\omega\left(x'\cos\phi+z'\sin\phi\right).$$

All terms of the form xy' or $y'y$ can be neglected because it is assumed x and y remain small. Also, the pendulum is assumed to be long with a heavy weight so that x' and y' are also small. With these simplifying assumptions, the equations of motion become

$$x''+g\frac{x}{l} = 2\omega y'\cos\phi$$

and

$$y''+g\frac{y}{l} = -2\omega x'\cos\phi.$$

[2] There is such a pendulum in the Eyring building at BYU and to keep people from touching it, there is a little sign which says Warning! 1000 ohms.

These equations are of the form

$$x'' + a^2 x = by',\ y'' + a^2 y = -bx' \tag{2.32}$$

where $a^2 = \frac{g}{l}$ and $b = 2\omega\cos\phi$. Then it is fairly tedious but routine to verify that for each constant, c,

$$x = c\sin\left(\frac{bt}{2}\right)\sin\left(\frac{\sqrt{b^2+4a^2}}{2}t\right),\ y = c\cos\left(\frac{bt}{2}\right)\sin\left(\frac{\sqrt{b^2+4a^2}}{2}t\right) \tag{2.33}$$

yields a solution to 2.32 along with the initial conditions,

$$x(0) = 0, y(0) = 0, x'(0) = 0, y'(0) = \frac{c\sqrt{b^2+4a^2}}{2}. \tag{2.34}$$

It is clear from experiments with the pendulum that the earth does indeed rotate out from under it causing the plane of vibration of the pendulum to appear to rotate. The purpose of this discussion is not to establish these self evident facts but to predict how long it takes for the plane of vibration to make one revolution. Therefore, there will be some instant in time at which the pendulum will be vibrating in a plane determined by **k** and **j**. (Recall **k** points away from the center of the earth and **j** points East.) At this instant in time, defined as $t = 0$, the conditions of 2.34 will hold for some value of c and so the solution to 2.32 having these initial conditions will be those of 2.33 by uniqueness of the initial value problem. Writing these solutions differently,

$$\begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = c\begin{pmatrix} \sin\left(\frac{bt}{2}\right) \\ \cos\left(\frac{bt}{2}\right) \end{pmatrix}\sin\left(\frac{\sqrt{b^2+4a^2}}{2}t\right)$$

This is very interesting! The vector, $c\begin{pmatrix} \sin\left(\frac{bt}{2}\right) \\ \cos\left(\frac{bt}{2}\right) \end{pmatrix}$ always has magnitude equal to $|c|$ but its direction changes very slowly because b is very small. The plane of vibration is determined by this vector and the vector **k**. The term $\sin\left(\frac{\sqrt{b^2+4a^2}}{2}t\right)$ changes relatively fast and takes values between -1 and 1. This is what describes the actual observed vibrations of the pendulum. Thus the plane of vibration will have made one complete revolution when $t = T$ for

$$\frac{bT}{2} \equiv 2\pi.$$

Therefore, the time it takes for the earth to turn out from under the pendulum is

$$T = \frac{4\pi}{2\omega\cos\phi} = \frac{2\pi}{\omega}\sec\phi.$$

Since ω is the angular speed of the rotating earth, it follows $\omega = \frac{2\pi}{24} = \frac{\pi}{12}$ in radians per hour. Therefore, the above formula implies

$$T = 24\sec\phi.$$

I think this is really amazing. You could actually determine latitude, not by taking readings with instruments using the North Star but by doing an experiment with a big pendulum. You would set it vibrating, observe T in hours, and then solve the above equation for ϕ. Also note the pendulum would not appear to change its plane of vibration at the equator because $\lim_{\phi\to\pi/2}\sec\phi = \infty$.

The Coriolis acceleration is also responsible for the phenomenon of the next example.

Example 2.6.7 *It is known that low pressure areas rotate counterclockwise as seen from above in the Northern hemisphere but clockwise in the Southern hemisphere. Why?*

Neglect accelerations other than the Coriolis acceleration and the following acceleration which comes from an assumption that the point $\mathbf{p}(t)$ is the location of the lowest pressure.

$$\mathbf{a} = -a(r_B)\,\mathbf{r}_B$$

where $r_B = r$ will denote the distance from the fixed point $\mathbf{p}(t)$ on the earth's surface which is also the lowest pressure point. Of course the situation could be more complicated but this will suffice to explain the above question. Then the acceleration observed by a person on the earth relative to the apparently fixed vectors, $\mathbf{i},\mathbf{k},\mathbf{j}$, is

$$\mathbf{a}_B = -a(r_B)(x\mathbf{i}+y\mathbf{j}+z\mathbf{k}) - 2\omega\left[-y'\cos(\phi)\,\mathbf{i}+(x'\cos(\phi)+z'\sin(\phi))\,\mathbf{j}-(y'\sin(\phi)\,\mathbf{k})\right]$$

Therefore, one obtains some differential equations from $\mathbf{a}_B = x''\mathbf{i}+y''\mathbf{j}+z''\mathbf{k}$ by matching the components. These are

$$\begin{aligned} x''+a(r_B)\,x &= 2\omega y'\cos\phi \\ y''+a(r_B)\,y &= -2\omega x'\cos\phi - 2\omega z'\sin(\phi) \\ z''+a(r_B)\,z &= 2\omega y'\sin\phi \end{aligned}$$

Now remember, the vectors, $\mathbf{i},\mathbf{j},\mathbf{k}$ are fixed relative to the earth and so are constant vectors. Therefore, from the properties of the determinant and the above differential equations,

$$(\mathbf{r}_B'\times\mathbf{r}_B)' = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ x' & y' & z' \\ x & y & z \end{vmatrix}' = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ x'' & y'' & z'' \\ x & y & z \end{vmatrix}$$

$$= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ -a(r_B)\,x+2\omega y'\cos\phi & -a(r_B)\,y-2\omega x'\cos\phi-2\omega z'\sin(\phi) & -a(r_B)\,z+2\omega y'\sin\phi \\ x & y & z \end{vmatrix}$$

Then the $\mathbf{k}^{th}$ component of this cross product equals

$$\omega\cos(\phi)\left(y^2+x^2\right)' + 2\omega x z'\sin(\phi).$$

The first term will be negative because it is assumed $\mathbf{p}(t)$ is the location of low pressure causing y^2+x^2 to be a decreasing function. If it is assumed there is not a substantial motion in the $\mathbf{k}$ direction, so that z is fairly constant and the last term can be neglected, then the $\mathbf{k}^{th}$ component of $(\mathbf{r}_B'\times\mathbf{r}_B)'$ is negative provided $\phi\in\left(0,\frac{\pi}{2}\right)$ and positive if $\phi\in\left(\frac{\pi}{2},\pi\right)$. Beginning with a point at rest, this implies $\mathbf{r}_B'\times\mathbf{r}_B=\mathbf{0}$ initially and then the above implies its $\mathbf{k}^{th}$ component is negative in the upper hemisphere when $\phi<\pi/2$ and positive in the lower hemisphere when $\phi>\pi/2$. Using the right hand and the geometric definition of the cross product, this shows clockwise rotation in the lower hemisphere and counter clockwise rotation in the upper hemisphere.

Note also that as ϕ gets close to $\pi/2$ near the equator, the above reasoning tends to break down because $\cos(\phi)$ becomes close to zero. Therefore, the motion towards the low pressure has to be more pronounced in comparison with the motion in the $\mathbf{k}$ direction in order to draw this conclusion.

2.7 Exercises

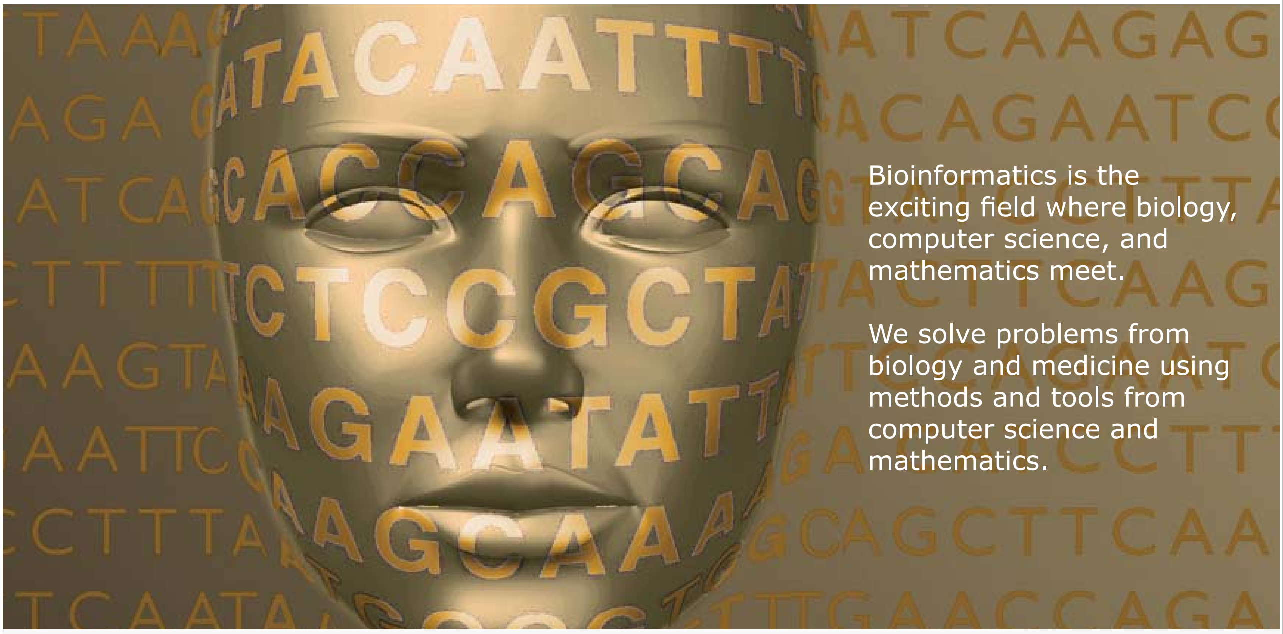
1. Show the map $T : \mathbb{R}^n \to \mathbb{R}^m$ defined by $T(\mathbf{x}) = A\mathbf{x}$ where A is an $m \times n$ matrix and $\mathbf{x}$ is an $m \times 1$ column vector is a linear transformation.

2. Find the matrix for the linear transformation which rotates every vector in $\mathbb{R}^2$ through an angle of $\pi/3$.

3. Find the matrix for the linear transformation which rotates every vector in $\mathbb{R}^2$ through an angle of $\pi/4$.

4. Find the matrix for the linear transformation which rotates every vector in $\mathbb{R}^2$ through an angle of $-\pi/3$.

5. Find the matrix for the linear transformation which rotates every vector in $\mathbb{R}^2$ through an angle of $2\pi/3$.



6. Find the matrix for the linear transformation which rotates every vector in $\mathbb{R}^2$ through an angle of $\pi/12$. **Hint:** Note that $\pi/12 = \pi/3 - \pi/4$.

7. Find the matrix for the linear transformation which rotates every vector in $\mathbb{R}^2$ through an angle of $2\pi/3$ and then reflects across the x axis.

8. Find the matrix for the linear transformation which rotates every vector in $\mathbb{R}^2$ through an angle of $\pi/3$ and then reflects across the x axis.

9. Find the matrix for the linear transformation which rotates every vector in $\mathbb{R}^2$ through an angle of $\pi/4$ and then reflects across the x axis.

10. Find the matrix for the linear transformation which rotates every vector in $\mathbb{R}^2$ through an angle of $\pi/6$ and then reflects across the x axis followed by a reflection across the y axis.

11. Find the matrix for the linear transformation which reflects every vector in $\mathbb{R}^2$ across the x axis and then rotates every vector through an angle of $\pi/4$.

12. Find the matrix for the linear transformation which rotates every vector in $\mathbb{R}^2$ through an angle of $\pi/4$ and next reflects every vector across the x axis. Compare with the above problem.

13. Find the matrix for the linear transformation which reflects every vector in $\mathbb{R}^2$ across the x axis and then rotates every vector through an angle of $\pi/6$.

14. Find the matrix for the linear transformation which reflects every vector in $\mathbb{R}^2$ across the y axis and then rotates every vector through an angle of $\pi/6$.

15. Find the matrix for the linear transformation which rotates every vector in $\mathbb{R}^2$ through an angle of $5\pi/12$. **Hint:** Note that $5\pi/12 = 2\pi/3 - \pi/4$.

16. Find the matrix for $\operatorname{proj}_{\mathbf{u}}(\mathbf{v})$ where $\mathbf{u} = (1, -2, 3)^T$.

17. Find the matrix for $\operatorname{proj}_{\mathbf{u}}(\mathbf{v})$ where $\mathbf{u} = (1, 5, 3)^T$.

18. Find the matrix for $\operatorname{proj}_{\mathbf{u}}(\mathbf{v})$ where $\mathbf{u} = (1, 0, 3)^T$.

19. Give an example of a 2×2 matrix A which has all its entries nonzero and satisfies $A^2 = A$. A matrix which satisfies $A^2 = A$ is called idempotent.

20. Let A be an $m \times n$ matrix and let B be an $n \times m$ matrix where $n < m$. Show that AB cannot have an inverse.

21. Find $\ker(A)$ for
$$A = \begin{pmatrix} 1 & 2 & 3 & 2 & 1 \\ 0 & 2 & 1 & 1 & 2 \\ 1 & 4 & 4 & 3 & 3 \\ 0 & 2 & 1 & 1 & 2 \end{pmatrix}.$$
Recall $\ker(A)$ is just the set of solutions to $A\mathbf{x} = \mathbf{0}$.

22. If A is a linear transformation, and $A\mathbf{x}_p = \mathbf{b}$, show that the general solution to the equation $A\mathbf{x} = \mathbf{b}$ is of the form $\mathbf{x}_p + \mathbf{y}$ where $\mathbf{y} \in \ker(A)$. By this I mean to show that whenever $A\mathbf{z} = \mathbf{b}$ there exists $\mathbf{y} \in \ker(A)$ such that $\mathbf{x}_p + \mathbf{y} = \mathbf{z}$. For the definition of $\ker(A)$ see Problem 21.

23. Using Problem 21, find the general solution to the following linear system.

$$\begin{pmatrix} 1 & 2 & 3 & 2 & 1 \\ 0 & 2 & 1 & 1 & 2 \\ 1 & 4 & 4 & 3 & 3 \\ 0 & 2 & 1 & 1 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} 11 \\ 7 \\ 18 \\ 7 \end{pmatrix}$$

24. Using Problem 21, find the general solution to the following linear system.

$$\begin{pmatrix} 1 & 2 & 3 & 2 & 1 \\ 0 & 2 & 1 & 1 & 2 \\ 1 & 4 & 4 & 3 & 3 \\ 0 & 2 & 1 & 1 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} 6 \\ 7 \\ 13 \\ 7 \end{pmatrix}$$

25. Show that the function $T_{\mathbf{u}}$ defined by $T_{\mathbf{u}}(\mathbf{v}) \equiv \mathbf{v} - \text{proj}_{\mathbf{u}}(\mathbf{v})$ is also a linear transformation.

26. If $\mathbf{u} = (1, 2, 3)^T$, as in Example 9.3.22 and $T_{\mathbf{u}}$ is given in the above problem, find the matrix $A_{\mathbf{u}}$ which satisfies $A_{\mathbf{u}}\mathbf{x} = T_{\mathbf{u}}(\mathbf{x})$.

27. Suppose V is a subspace of $\mathbb{F}^n$ and $T : V \to \mathbb{F}^p$ is a nonzero linear transformation. Show that there exists a basis for $\text{Im}(T) \equiv T(V)$

$$\{T\mathbf{v}_1, \cdots, T\mathbf{v}_m\}$$

and that in this situation,

$$\{\mathbf{v}_1, \cdots, \mathbf{v}_m\}$$

is linearly independent.

28. $\uparrow$In the situation of Problem 27 where V is a subspace of $\mathbb{F}^n$, show that there exists $\{\mathbf{z}_1, \cdots, \mathbf{z}_r\}$ a basis for $\ker(T)$. (Recall Theorem 2.4.12. Since $\ker(T)$ is a subspace, it has a basis.) Now for an arbitrary $T\mathbf{v} \in T(V)$, explain why

$$T\mathbf{v} = a_1 T\mathbf{v}_1 + \cdots + a_m T\mathbf{v}_m$$

and why this implies

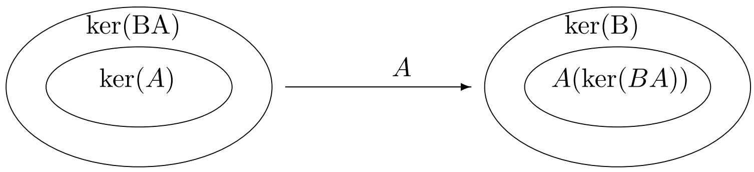
$$\mathbf{v} - (a_1\mathbf{v}_1 + \cdots + a_m\mathbf{v}_m) \in \ker(T).$$

Then explain why $V = \text{span}(\mathbf{v}_1, \cdots, \mathbf{v}_m, \mathbf{z}_1, \cdots, \mathbf{z}_r)$.

29. $\uparrow$In the situation of the above problem, show $\{\mathbf{v}_1, \cdots, \mathbf{v}_m, \mathbf{z}_1, \cdots, \mathbf{z}_r\}$ is a basis for V and therefore, $\dim(V) = \dim(\ker(T)) + \dim(T(V))$.

30. $\uparrow$Let A be a linear transformation from V to W and let B be a linear transformation from W to U where V, W, U are all subspaces of some $\mathbb{F}^p$. Explain why

$$A(\ker(BA)) \subseteq \ker(B), \ \ker(A) \subseteq \ker(BA).$$



31. $\uparrow$Let $\{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$ be a basis of $\ker(A)$ and let $\{A\mathbf{y}_1, \cdots, A\mathbf{y}_m\}$ be a basis of $A(\ker(BA))$. Let $\mathbf{z} \in \ker(BA)$. Explain why

$$Az \in \operatorname{span}\{A\mathbf{y}_1, \cdots, A\mathbf{y}_m\}$$

and why there exist scalars a_i such that

$$A(z - (a_1\mathbf{y}_1 + \cdots + a_m\mathbf{y}_m)) = 0$$

and why it follows $z - (a_1\mathbf{y}_1 + \cdots + a_m\mathbf{y}_m) \in \operatorname{span}\{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$. Now explain why

$$\ker(BA) \subseteq \operatorname{span}\{\mathbf{x}_1, \cdots, \mathbf{x}_n, \mathbf{y}_1, \cdots, \mathbf{y}_m\}$$

and so

$$\dim(\ker(BA)) \leq \dim(\ker(B)) + \dim(\ker(A)).$$

This important inequality is due to Sylvester. Show that equality holds if and only if $A(\ker BA) = \ker(B)$.

32. Generalize the result of the previous problem to any finite product of linear mappings.

33. If $W \subseteq V$ for W, V two subspaces of $\mathbb{F}^n$ and if $\dim(W) = \dim(V)$, show $W = V$.

34. Let V be a subspace of $\mathbb{F}^n$and let $V_1, \cdots, V_m$ be subspaces, each contained in V. Then

$$V = V_1 \oplus \cdots \oplus V_m \tag{2.35}$$

if every $v \in V$ can be written in a unique way in the form

$$v = v_1 + \cdots + v_m$$

where each $v_i \in V_i$. This is called a direct sum. If this uniqueness condition does not hold, then one writes

$$V = V_1 + \cdots + V_m$$

and this symbol means all vectors of the form

$$v_1 + \cdots + v_m, \ v_j \in V_j \text{ for each } j.$$

Show 2.35 is equivalent to saying that if

$$0 = v_1 + \cdots + v_m, \ v_j \in V_j \text{ for each } j,$$

then each $v_j = 0$. Next show that in the situation of 2.35, if $\beta_i = \left\{u_1^i, \cdots, u_{m_i}^i\right\}$ is a basis for V_i, then $\{\beta_1, \cdots, \beta_m\}$ is a basis for V.

35. ↑Suppose you have finitely many linear mappings $L_1, L_2, \cdots, L_m$ which map V to V where V is a subspace of $\mathbb{F}^n$ and suppose they commute. That is, $L_i L_j = L_j L_i$ for all i, j. Also suppose L_k is one to one on $\ker(L_j)$ whenever $j \neq k$. Letting P denote the product of these linear transformations, $P = L_1 L_2 \cdots L_m$, first show

$$\ker(L_1) + \cdots + \ker(L_m) \subseteq \ker(P)$$

Next show $L_j : \ker(L_i) \to \ker(L_i)$. Then show

$$\ker(L_1) + \cdots + \ker(L_m) = \ker(L_1) \oplus \cdots \oplus \ker(L_m).$$

Using Sylvester's theorem, and the result of Problem 33, show

$$\ker(P) = \ker(L_1) \oplus \cdots \oplus \ker(L_m)$$

Hint: By Sylvester's theorem and the above problem,

$$\begin{aligned} \dim(\ker(P)) &\leq \sum_i \dim(\ker(L_i)) \\ &= \dim(\ker(L_1) \oplus \cdots \oplus \ker(L_m)) \leq \dim(\ker(P)) \end{aligned}$$

Now consider Problem 33.

36. Let $\mathcal{M}(\mathbb{F}^n, \mathbb{F}^n)$ denote the set of all $n \times n$ matrices having entries in $\mathbb{F}$. With the usual operations of matrix addition and scalar multiplications, explain why $\mathcal{M}(\mathbb{F}^n, \mathbb{F}^n)$ can be considered as $\mathbb{F}^{n^2}$. Give a basis for $\mathcal{M}(\mathbb{F}^n, \mathbb{F}^n)$. If $A \in \mathcal{M}(\mathbb{F}^n, \mathbb{F}^n)$, explain why there exists a monic (leading coefficient equals 1) polynomial of the form

$$\lambda^k + a_{k-1}\lambda^{k-1} + \cdots + a_1\lambda + a_0$$

such that

$$A^k + a_{k-1}A^{k-1} + \cdots + a_1 A + a_0 I = 0$$

The minimal polynomial of A is the polynomial like the above, for which $p(A) = 0$ which has smallest degree. I will discuss the uniqueness of this polynomial later. **Hint:** Consider the matrices $I, A, A^2, \cdots, A^{n^2}$. There are $n^2 + 1$ of these matrices. Can they be linearly independent? Now consider all polynomials and pick one of smallest degree and then divide by the leading coefficient.

37. $\uparrow$Suppose the field of scalars is $\mathbb{C}$ and A is an $n \times n$ matrix. From the preceding problem, and the fundamental theorem of algebra, this minimal polynomial factors

$$(\lambda - \lambda_1)^{r_1} (\lambda - \lambda_2)^{r_2} \cdots (\lambda - \lambda_k)^{r_k}$$

where r_j is the algebraic multiplicity of λ_j, and the λ_j are distinct. Thus

$$(A - \lambda_1 I)^{r_1} (A - \lambda_2 I)^{r_2} \cdots (A - \lambda_k I)^{r_k} = 0$$

and so, letting $P = (A - \lambda_1 I)^{r_1} (A - \lambda_2 I)^{r_2} \cdots (A - \lambda_k I)^{r_k}$ and $L_j = (A - \lambda_j I)^{r_j}$ apply the result of Problem 35 to verify that

$$\mathbb{C}^n = \ker(L_1) \oplus \cdots \oplus \ker(L_k)$$

and that $A : \ker(L_j) \to \ker(L_j)$. In this context, $\ker(L_j)$ is called the generalized eigenspace for λ_j. You need to verify the conditions of the result of this problem hold.

38. In the context of Problem 37, show there exists a nonzero vector $\mathbf{x}$ such that

$$(A - \lambda_j I)\,\mathbf{x} = \mathbf{0}.$$

This is called an eigenvector and the λ_j is called an eigenvalue. **Hint:**There must exist a vector $\mathbf{y}$ such that

$$(A - \lambda_1 I)^{r_1} (A - \lambda_2 I)^{r_2} \cdots (A - \lambda_j I)^{r_j - 1} \cdots (A - \lambda_k I)^{r_k}\,\mathbf{y} = \mathbf{z} \neq \mathbf{0}$$

Why? Now what happens if you do $(A - \lambda_j I)$ to $\mathbf{z}$?

39. Suppose $Q(t)$ is an orthogonal matrix. This means $Q(t)$ is a real $n \times n$ matrix which satisfies

$$Q(t)\, Q(t)^T = I$$

Suppose also the entries of $Q(t)$ are differentiable. Show $\left(Q^T\right)' = -Q^T Q' Q^T$.

40. Remember the Coriolis force was $2\boldsymbol{\Omega} \times \mathbf{v}_B$ where $\boldsymbol{\Omega}$ was a particular vector which came from the matrix $Q(t)$ as described above. Show that

$$Q(t) = \begin{pmatrix} \mathbf{i}(t)\cdot\mathbf{i}(t_0) & \mathbf{j}(t)\cdot\mathbf{i}(t_0) & \mathbf{k}(t)\cdot\mathbf{i}(t_0) \\ \mathbf{i}(t)\cdot\mathbf{j}(t_0) & \mathbf{j}(t)\cdot\mathbf{j}(t_0) & \mathbf{k}(t)\cdot\mathbf{j}(t_0) \\ \mathbf{i}(t)\cdot\mathbf{k}(t_0) & \mathbf{j}(t)\cdot\mathbf{k}(t_0) & \mathbf{k}(t)\cdot\mathbf{k}(t_0) \end{pmatrix}.$$

There will be no Coriolis force exactly when $\boldsymbol{\Omega} = \mathbf{0}$ which corresponds to $Q'(t) = 0$. When will $Q'(t) = 0$?

41. An illustration used in many beginning physics books is that of firing a rifle horizontally and dropping an identical bullet from the same height above the perfectly flat ground followed by an assertion that the two bullets will hit the ground at exactly the same time. Is this true on the rotating earth assuming the experiment takes place over a large perfectly flat field so the curvature of the earth is not an issue? Explain. What other irregularities will occur? Recall the Coriolis acceleration is $2\omega\left[(-y'\cos\phi)\,\mathbf{i} + (x'\cos\phi + z'\sin\phi)\,\mathbf{j} - (y'\sin\phi)\,\mathbf{k}\right]$ where $\mathbf{k}$ points away from the center of the earth, $\mathbf{j}$ points East, and $\mathbf{i}$ points South.

Determinants

3.1 Basic Techniques And Properties

Let A be an $n \times n$ matrix. The determinant of A, denoted as $\det(A)$ is a number. If the matrix is a 2×2 matrix, this number is very easy to find.

Definition 3.1.1 *Let* $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. *Then*

$$\det(A) \equiv ad - cb.$$

The determinant is also often denoted by enclosing the matrix with two vertical lines. Thus

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix}.$$

Example 3.1.2 *Find* $\det \begin{pmatrix} 2 & 4 \\ -1 & 6 \end{pmatrix}$.

From the definition this is just $(2)(6) - (-1)(4) = 16$.

Assuming the determinant has been defined for $k \times k$ matrices for $k \leq n-1$, it is now time to define it for $n \times n$ matrices.

Definition 3.1.3 *Let $A = (a_{ij})$ be an $n \times n$ matrix. Then a new matrix called the cofactor matrix, $\operatorname{cof}(A)$ is defined by $\operatorname{cof}(A) = (c_{ij})$ where to obtain c_{ij} delete the i^{th} row and the j^{th} column of A, take the determinant of the $(n-1) \times (n-1)$ matrix which results, (This is called the ij^{th} minor of A.) and then multiply this number by $(-1)^{i+j}$. To make the formulas easier to remember, $\operatorname{cof}(A)_{ij}$ will denote the ij^{th} entry of the cofactor matrix.*

Now here is the definition of the determinant given recursively.

Theorem 3.1.4 *Let A be an $n \times n$ matrix where $n \geq 2$. Then*

$$\det(A) = \sum_{j=1}^{n} a_{ij} \operatorname{cof}(A)_{ij} = \sum_{i=1}^{n} a_{ij} \operatorname{cof}(A)_{ij}. \tag{3.1}$$

The first formula consists of expanding the determinant along the i^{th} row and the second expands the determinant along the j^{th} column.

Note that for a $n \times n$ matrix, you will need $n!$ terms to evaluate the determinant in this way. If $n = 10$, this is $10! = 3,628\,,800$ terms. This is a lot of terms.

In addition to the difficulties just discussed, why is the determinant well defined? Why should you get the same thing when you expand along any row or column? I think you should regard this claim that you always get the same answer by picking any row or column with considerable skepticism. It is incredible and not at all obvious. However, it requires a little effort to establish it. This is done in the section on the theory of the determinant which follows.

Notwithstanding the difficulties involved in using the method of Laplace expansion, certain types of matrices are very easy to deal with.

Definition 3.1.5 *A matrix M, is upper triangular if $M_{ij} = 0$ whenever $i > j$. Thus such a matrix equals zero below the main diagonal, the entries of the form M_{ii}, as shown.*

$$\begin{pmatrix} * & * & \cdots & * \\ 0 & * & \ddots & \vdots \\ \vdots & \ddots & \ddots & * \\ 0 & \cdots & 0 & * \end{pmatrix}$$

A lower triangular matrix is defined similarly as a matrix for which all entries above the main diagonal are equal to zero.

You should verify the following using the above theorem on Laplace expansion.

Corollary 3.1.6 *Let M be an upper (lower) triangular matrix. Then $\det(M)$ is obtained by taking the product of the entries on the main diagonal.*

Proof: The corollary is true if the matrix is one to one. Suppose it is $n \times n$. Then the matrix is of the form

$$\begin{pmatrix} m_{11} & \mathbf{a} \\ \mathbf{0} & M_1 \end{pmatrix}$$

where M_1 is $(n-1) \times (n-1)$. Then expanding along the first row, you get $m_{11} \det(M_1) + 0$. Then use the induction hypothesis to obtain that $\det(M_1) = \prod_{i=2}^{n} m_{ii}$. ■

Example 3.1.7 *Let*

$$A = \begin{pmatrix} 1 & 2 & 3 & 77 \\ 0 & 2 & 6 & 7 \\ 0 & 0 & 3 & 33.7 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

Find $\det(A)$.

From the above corollary, this is -6.

There are many properties satisfied by determinants. Some of the most important are listed in the following theorem.

Theorem 3.1.8 *If two rows or two columns in an $n \times n$ matrix A are switched, the determinant of the resulting matrix equals (-1) times the determinant of the original matrix. If A is an $n \times n$ matrix in which two rows are equal or two columns are equal then $\det(A) = 0$. Suppose the i^{th} row of A equals $(xa_1 + yb_1, \cdots, xa_n + yb_n)$. Then*

$$\det(A) = x \det(A_1) + y \det(A_2)$$

where the i^{th} row of A_1 is $(a_1, \cdots, a_n)$ and the i^{th} row of A_2 is $(b_1, \cdots, b_n)$, all other rows of A_1 and A_2 coinciding with those of A. In other words, det is a linear function of each row A. The same is true with the word "row" replaced with the word "column". In addition to this, if A and B are $n \times n$ matrices, then

$$\det(AB) = \det(A) \det(B),$$

and if A is an $n \times n$ matrix, then

$$\det(A) = \det\left(A^T\right).$$

This theorem implies the following corollary which gives a way to find determinants. As I pointed out above, the method of Laplace expansion will not be practical for any matrix of large size.

Corollary 3.1.9 *Let A be an $n \times n$ matrix and let B be the matrix obtained by replacing the i^{th} row (column) of A with the sum of the i^{th} row (column) added to a multiple of another row (column). Then* $\det(A) = \det(B)$. *If B is the matrix obtained from A be replacing the i^{th} row (column) of A by a times the i^{th} row (column) then* $a\det(A) = \det(B)$.

Here is an example which shows how to use this corollary to find a determinant.

Example 3.1.10 *Find the determinant of the matrix*

$$A = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 1 & 2 & 3 \\ 4 & 5 & 4 & 3 \\ 2 & 2 & -4 & 5 \end{pmatrix}$$

Replace the second row by (-5) times the first row added to it. Then replace the third row by (-4) times the first row added to it. Finally, replace the fourth row by (-2) times the first row added to it. This yields the matrix

$$B = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & -9 & -13 & -17 \\ 0 & -3 & -8 & -13 \\ 0 & -2 & -10 & -3 \end{pmatrix}$$

and from the above corollary, it has the same determinant as A. Now using the corollary some more, $\det(B) = \left(\frac{-1}{3}\right)\det(C)$ where

$$C = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & 0 & 11 & 22 \\ 0 & -3 & -8 & -13 \\ 0 & 6 & 30 & 9 \end{pmatrix}.$$

The second row was replaced by (-3) times the third row added to the second row and then the last row was multiplied by (-3). Now replace the last row with 2 times the third added to it and then switch the third and second rows. Then $\det(C) = -\det(D)$ where

$$D = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & -3 & -8 & -13 \\ 0 & 0 & 11 & 22 \\ 0 & 0 & 14 & -17 \end{pmatrix}$$

You could do more row operations or you could note that this can be easily expanded along the first column followed by expanding the 3×3 matrix which results along its first column. Thus

$$\det(D) = 1(-3)\begin{vmatrix} 11 & 22 \\ 14 & -17 \end{vmatrix} = 1485$$

and so $\det(C) = -1485$ and $\det(A) = \det(B) = \left(\frac{-1}{3}\right)(-1485) = 495$.

The theorem about expanding a matrix along any row or column also provides a way to give a formula for the inverse of a matrix. Recall the definition of the inverse of a matrix in Definition 2.1.22 on Page 61. The following theorem gives a formula for the inverse of a matrix. It is proved in the next section.

Theorem 3.1.11 *A^{-1} exists if and only if $\det(A) \neq 0$. If $\det(A) \neq 0$, then $A^{-1} = \left(a_{ij}^{-1}\right)$ where*

$$a_{ij}^{-1} = \det(A)^{-1} \operatorname{cof}(A)_{ji}$$

for $\operatorname{cof}(A)_{ij}$ the ij^{th} cofactor of A.

Theorem 3.1.11 says that to find the inverse, take the transpose of the cofactor matrix and divide by the determinant. The transpose of the cofactor matrix is called the adjugate or sometimes the classical adjoint of the matrix A. It is an abomination to call it the adjoint although you do sometimes see it referred to in this way. In words, A^{-1} is equal to one over the determinant of A times the adjugate matrix of A.

Example 3.1.12 *Find the inverse of the matrix*

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 0 & 1 \\ 1 & 2 & 1 \end{pmatrix}$$

First find the determinant of this matrix. This is seen to be 12. The cofactor matrix of A is

$$\begin{pmatrix} -2 & -2 & 6 \\ 4 & -2 & 0 \\ 2 & 8 & -6 \end{pmatrix}.$$

Each entry of A was replaced by its cofactor. Therefore, from the above theorem, the inverse of A should equal

$$\frac{1}{12}\begin{pmatrix} -2 & -2 & 6 \\ 4 & -2 & 0 \\ 2 & 8 & -6 \end{pmatrix}^T = \begin{pmatrix} -\frac{1}{6} & \frac{1}{3} & \frac{1}{6} \\ -\frac{1}{6} & -\frac{1}{6} & \frac{2}{3} \\ \frac{1}{2} & 0 & -\frac{1}{2} \end{pmatrix}.$$

This way of finding inverses is especially useful in the case where it is desired to find the inverse of a matrix whose entries are functions.

Example 3.1.13 *Suppose*

$$A(t) = \begin{pmatrix} e^t & 0 & 0 \\ 0 & \cos t & \sin t \\ 0 & -\sin t & \cos t \end{pmatrix}$$

Find $A(t)^{-1}$.

First note $\det(A(t)) = e^t$. A routine computation using the above theorem shows that this inverse is

$$\frac{1}{e^t}\begin{pmatrix} 1 & 0 & 0 \\ 0 & e^t\cos t & e^t\sin t \\ 0 & -e^t\sin t & e^t\cos t \end{pmatrix}^T = \begin{pmatrix} e^{-t} & 0 & 0 \\ 0 & \cos t & -\sin t \\ 0 & \sin t & \cos t \end{pmatrix}.$$

This formula for the inverse also implies a famous procedure known as Cramer's rule. Cramer's rule gives a formula for the solutions, $\mathbf{x}$, to a system of equations, $A\mathbf{x} = \mathbf{y}$.

In case you are solving a system of equations, $A\mathbf{x} = \mathbf{y}$ for $\mathbf{x}$, it follows that if A^{-1} exists,

$$\mathbf{x} = \left(A^{-1}A\right)\mathbf{x} = A^{-1}\left(A\mathbf{x}\right) = A^{-1}\mathbf{y}$$

thus solving the system. Now in the case that A^{-1} exists, there is a formula for A^{-1} given above. Using this formula,

$$x_i = \sum_{j=1}^{n} a_{ij}^{-1} y_j = \sum_{j=1}^{n} \frac{1}{\det(A)} \operatorname{cof}(A)_{ji} y_j.$$

By the formula for the expansion of a determinant along a column,

$$x_i = \frac{1}{\det(A)} \det \begin{pmatrix} * & \cdots & y_1 & \cdots & * \\ \vdots & & \vdots & & \vdots \\ * & \cdots & y_n & \cdots & * \end{pmatrix},$$

where here the i^{th} column of A is replaced with the column vector, $(y_1 \cdots, y_n)^T$, and the determinant of this modified matrix is taken and divided by $\det(A)$. This formula is known as Cramer's rule.

Procedure 3.1.14 *Suppose A is an $n \times n$ matrix and it is desired to solve the system $A\mathbf{x} = \mathbf{y}, \mathbf{y} = (y_1, \cdots, y_n)^T$ for $\mathbf{x} = (x_1, \cdots, x_n)^T$. Then Cramer's rule says*

$$x_i = \frac{\det A_i}{\det A}$$

where A_i is obtained from A by replacing the i^{th} column of A with the column $(y_1, \cdots, y_n)^T$.

The following theorem is of fundamental importance and ties together many of the ideas presented above. It is proved in the next section.

Theorem 3.1.15 *Let A be an $n \times n$ matrix. Then the following are equivalent.*

1. *A is one to one.*
2. *A is onto.*
3. $\det(A) \neq 0$.

3.2 Exercises

1. Find the determinants of the following matrices.

   (a) $\begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 2 \\ 0 & 9 & 8 \end{pmatrix}$ (The answer is 31.)

   (b) $\begin{pmatrix} 4 & 3 & 2 \\ 1 & 7 & 8 \\ 3 & -9 & 3 \end{pmatrix}$ (The answer is 375.)

(c) $\begin{pmatrix} 1 & 2 & 3 & 2 \\ 1 & 3 & 2 & 3 \\ 4 & 1 & 5 & 0 \\ 1 & 2 & 1 & 2 \end{pmatrix}$, (The answer is -2.)

2. If A^{-1} exist, what is the relationship between $\det(A)$ and $\det\left(A^{-1}\right)$. Explain your answer.

3. Let A be an $n \times n$ matrix where n is odd. Suppose also that A is skew symmetric. This means $A^T = -A$. Show that $\det(A) = 0$.

4. Is it true that $\det(A + B) = \det(A) + \det(B)$? If this is so, explain why it is so and if it is not so, give a counter example.

5. Let A be an $r \times r$ matrix and suppose there are $r-1$ rows (columns) such that all rows (columns) are linear combinations of these $r-1$ rows (columns). Show $\det(A) = 0$.

6. Show $\det(aA) = a^n \det(A)$ where here A is an $n \times n$ matrix and a is a scalar.

7. Suppose A is an upper triangular matrix. Show that A^{-1} exists if and only if all elements of the main diagonal are non zero. Is it true that A^{-1} will also be upper triangular? Explain. Is everything the same for lower triangular matrices?

8. Let A and B be two $n \times n$ matrices. $A \sim B$ (A is similar to B) means there exists an invertible matrix S such that $A = S^{-1}BS$. Show that if $A \sim B$, then $B \sim A$. Show also that $A \sim A$ and that if $A \sim B$ and $B \sim C$, then $A \sim C$.

9. In the context of Problem 8 show that if $A \sim B$, then $\det(A) = \det(B)$.

10. Let A be an $n \times n$ matrix and let $\mathbf{x}$ be a nonzero vector such that $A\mathbf{x} = \lambda\mathbf{x}$ for some scalar, λ. When this occurs, the vector, $\mathbf{x}$ is called an eigenvector and the scalar, λ is called an eigenvalue. It turns out that not every number is an eigenvalue. Only certain ones are. Why? **Hint:** Show that if $A\mathbf{x} = \lambda\mathbf{x}$, then $(\lambda I - A)\mathbf{x} = \mathbf{0}$. Explain why this shows that $(\lambda I - A)$ is not one to one and not onto. Now use Theorem 3.1.15 to argue $\det(\lambda I - A) = 0$. What sort of equation is this? How many solutions does it have?

11. Suppose $\det(\lambda I - A) = 0$. Show using Theorem 3.1.15 there exists $\mathbf{x} \neq \mathbf{0}$ such that $(\lambda I - A)\mathbf{x} = \mathbf{0}$.

12. Let $F(t) = \det\begin{pmatrix} a(t) & b(t) \\ c(t) & d(t) \end{pmatrix}$. Verify

$$F'(t) = \det\begin{pmatrix} a'(t) & b'(t) \\ c(t) & d(t) \end{pmatrix} + \det\begin{pmatrix} a(t) & b(t) \\ c'(t) & d'(t) \end{pmatrix}.$$

Now suppose

$$F(t) = \det\begin{pmatrix} a(t) & b(t) & c(t) \\ d(t) & e(t) & f(t) \\ g(t) & h(t) & i(t) \end{pmatrix}.$$

Use Laplace expansion and the first part to verify $F'(t) =$

$$\det\begin{pmatrix} a'(t) & b'(t) & c'(t) \\ d(t) & e(t) & f(t) \\ g(t) & h(t) & i(t) \end{pmatrix} + \det\begin{pmatrix} a(t) & b(t) & c(t) \\ d'(t) & e'(t) & f'(t) \\ g(t) & h(t) & i(t) \end{pmatrix}$$
$$+\det\begin{pmatrix} a(t) & b(t) & c(t) \\ d(t) & e(t) & f(t) \\ g'(t) & h'(t) & i'(t) \end{pmatrix}.$$

Conjecture a general result valid for $n \times n$ matrices and explain why it will be true. Can a similar thing be done with the columns?

13. Use the formula for the inverse in terms of the cofactor matrix to find the inverse of the matrix
$$A = \begin{pmatrix} e^t & 0 & 0 \\ 0 & e^t \cos t & e^t \sin t \\ 0 & e^t \cos t - e^t \sin t & e^t \cos t + e^t \sin t \end{pmatrix}.$$

14. Let A be an $r \times r$ matrix and let B be an $m \times m$ matrix such that $r + m = n$. Consider the following $n \times n$ block matrix
$$C = \begin{pmatrix} A & 0 \\ D & B \end{pmatrix}.$$
where the D is an $m \times r$ matrix, and the 0 is a $r \times m$ matrix. Letting I_k denote the $k \times k$ identity matrix, tell why
$$C = \begin{pmatrix} A & 0 \\ D & I_m \end{pmatrix} \begin{pmatrix} I_r & 0 \\ 0 & B \end{pmatrix}.$$
Now explain why $\det(C) = \det(A)\det(B)$. **Hint:** Part of this will require an explanation of why
$$\det \begin{pmatrix} A & 0 \\ D & I_m \end{pmatrix} = \det(A).$$
See Corollary 3.1.9.

15. Suppose Q is an orthogonal matrix. This means Q is a real $n \times n$ matrix which satisfies
$$QQ^T = I$$
Find the possible values for $\det(Q)$.

16. Suppose $Q(t)$ is an orthogonal matrix. This means $Q(t)$ is a real $n \times n$ matrix which satisfies
$$Q(t) Q(t)^T = I$$
Suppose $Q(t)$ is continuous for $t \in [a, b]$, some interval. Also suppose $\det(Q(t)) = 1$. Show that it follows $\det(Q(t)) = 1$ for all $t \in [a, b]$.

3.3 The Mathematical Theory Of Determinants

It is easiest to give a different definition of the determinant which is clearly well defined and then prove the earlier one in terms of Laplace expansion. Let $(i_1, \cdots, i_n)$ be an ordered list of numbers from $\{1, \cdots, n\}$. This means the order is important so $(1, 2, 3)$ and $(2, 1, 3)$ are different. There will be some repetition between this section and the earlier section on determinants. The main purpose is to give all the missing proofs. Two books which give a good introduction to determinants are Apostol [1] and Rudin [22]. A recent book which also has a good introduction is Baker [3]

3.3.1 The Function sgn

The following Lemma will be essential in the definition of the determinant.

Lemma 3.3.1 *There exists a unique function,* sgn_n *which maps each ordered list of numbers from* $\{1, \cdots, n\}$ *to one of the three numbers,* $0, 1$, *or* -1 *which also has the following properties.*

$$\operatorname{sgn}_n (1, \cdots, n) = 1 \tag{3.2}$$

$$\operatorname{sgn}_n (i_1, \cdots, p, \cdots, q, \cdots, i_n) = -\operatorname{sgn}_n (i_1, \cdots, q, \cdots, p, \cdots, i_n) \tag{3.3}$$

In words, the second property states that if two of the numbers are switched, the value of the function is multiplied by -1. *Also, in the case where* $n > 1$ *and* $\{i_1, \cdots, i_n\} = \{1, \cdots, n\}$ *so that every number from* $\{1, \cdots, n\}$ *appears in the ordered list,* $(i_1, \cdots, i_n)$,

$$\operatorname{sgn}_n (i_1, \cdots, i_{\theta-1}, n, i_{\theta+1}, \cdots, i_n) \equiv$$

$$(-1)^{n-\theta} \operatorname{sgn}_{n-1} (i_1, \cdots, i_{\theta-1}, i_{\theta+1}, \cdots, i_n) \tag{3.4}$$

where $n = i_\theta$ *in the ordered list,* $(i_1, \cdots, i_n)$.

Proof: To begin with, it is necessary to show the existence of such a function. This is clearly true if $n = 1$. Define $\operatorname{sgn}_1 (1) \equiv 1$ and observe that it works. No switching is possible. In the case where $n = 2$, it is also clearly true. Let $\operatorname{sgn}_2 (1, 2) = 1$ and $\operatorname{sgn}_2 (2, 1) = -1$ while $\operatorname{sgn}_2 (2, 2) = \operatorname{sgn}_2 (1, 1) = 0$ and verify it works. Assuming such a function exists for n, sgn_{n+1} will be defined in terms of sgn_n. If there are any repeated numbers in $(i_1, \cdots, i_{n+1})$, $\operatorname{sgn}_{n+1} (i_1, \cdots, i_{n+1}) \equiv 0$. If there are no repeats, then $n + 1$ appears somewhere in the ordered list. Let θ be the position of the number $n + 1$ in the list. Thus, the list is of the form $(i_1, \cdots, i_{\theta-1}, n + 1, i_{\theta+1}, \cdots, i_{n+1})$. From 3.4 it must be that

$$\operatorname{sgn}_{n+1} (i_1, \cdots, i_{\theta-1}, n + 1, i_{\theta+1}, \cdots, i_{n+1}) \equiv$$

$$(-1)^{n+1-\theta} \operatorname{sgn}_n (i_1, \cdots, i_{\theta-1}, i_{\theta+1}, \cdots, i_{n+1}).$$

It is necessary to verify this satisfies 3.2 and 3.3 with n replaced with $n + 1$. The first of these is obviously true because

$$\operatorname{sgn}_{n+1} (1, \cdots, n, n + 1) \equiv (-1)^{n+1-(n+1)} \operatorname{sgn}_n (1, \cdots, n) = 1.$$

If there are repeated numbers in $(i_1, \cdots, i_{n+1})$, then it is obvious 3.3 holds because both sides would equal zero from the above definition. It remains to verify 3.3 in the case where there are no numbers repeated in $(i_1, \cdots, i_{n+1})$. Consider

$$\operatorname{sgn}_{n+1} \left(i_1, \cdots, \overset{r}{p}, \cdots, \overset{s}{q}, \cdots, i_{n+1} \right),$$

where the r above the p indicates the number p is in the r^{th} position and the s above the q indicates that the number, q is in the s^{th} position. Suppose first that $r < \theta < s$. Then

$$\operatorname{sgn}_{n+1} \left(i_1, \cdots, \overset{r}{p}, \cdots, n \overset{\theta}{+} 1, \cdots, \overset{s}{q}, \cdots, i_{n+1} \right) \equiv$$

$$(-1)^{n+1-\theta} \operatorname{sgn}_n \left(i_1, \cdots, \overset{r}{p}, \cdots, \overset{s-1}{q}, \cdots, i_{n+1} \right)$$

while

$$\operatorname{sgn}_{n+1} \left(i_1, \cdots, \overset{r}{q}, \cdots, n \overset{\theta}{+} 1, \cdots, \overset{s}{p}, \cdots, i_{n+1} \right) \equiv$$

$$(-1)^{n+1-\theta} \operatorname{sgn}_n \left(i_1, \cdots, \overset{r}{q}, \cdots, \overset{s-1}{p}, \cdots, i_{n+1} \right)$$

and so, by induction, a switch of p and q introduces a minus sign in the result. Similarly, if $\theta > s$ or if $\theta < r$ it also follows that 3.3 holds. The interesting case is when $\theta = r$ or $\theta = s$. Consider the case where $\theta = r$ and note the other case is entirely similar.

$$\operatorname{sgn}_{n+1}\left(i_1, \cdots, \overset{r}{n+1}, \cdots, \overset{s}{q}, \cdots, i_{n+1}\right) \equiv$$
$$(-1)^{n+1-r} \operatorname{sgn}_n\left(i_1, \cdots, \overset{s-1}{q}, \cdots, i_{n+1}\right) \tag{3.5}$$

while

$$\operatorname{sgn}_{n+1}\left(i_1, \cdots, \overset{r}{q}, \cdots, \overset{s}{n+1}, \cdots, i_{n+1}\right) =$$
$$(-1)^{n+1-s} \operatorname{sgn}_n\left(i_1, \cdots, \overset{r}{q}, \cdots, i_{n+1}\right). \tag{3.6}$$

By making $s-1-r$ switches, move the q which is in the $s-1^{th}$ position in 3.5 to the r^{th} position in 3.6. By induction, each of these switches introduces a factor of -1 and so

$$\operatorname{sgn}_n\left(i_1, \cdots, \overset{s-1}{q}, \cdots, i_{n+1}\right) = (-1)^{s-1-r} \operatorname{sgn}_n\left(i_1, \cdots, \overset{r}{q}, \cdots, i_{n+1}\right).$$

Therefore,

$$\operatorname{sgn}_{n+1}\left(i_1, \cdots, \overset{r}{n+1}, \cdots, \overset{s}{q}, \cdots, i_{n+1}\right) = (-1)^{n+1-r} \operatorname{sgn}_n\left(i_1, \cdots, \overset{s-1}{q}, \cdots, i_{n+1}\right)$$
$$= (-1)^{n+1-r} (-1)^{s-1-r} \operatorname{sgn}_n\left(i_1, \cdots, \overset{r}{q}, \cdots, i_{n+1}\right)$$
$$= (-1)^{n+s} \operatorname{sgn}_n\left(i_1, \cdots, \overset{r}{q}, \cdots, i_{n+1}\right) = (-1)^{2s-1} (-1)^{n+1-s} \operatorname{sgn}_n\left(i_1, \cdots, \overset{r}{q}, \cdots, i_{n+1}\right)$$
$$= -\operatorname{sgn}_{n+1}\left(i_1, \cdots, \overset{r}{q}, \cdots, \overset{s}{n+1}, \cdots, i_{n+1}\right).$$

This proves the existence of the desired function. Uniqueness follows easily from the following lemma.

Lemma 3.3.2 *Every ordered list of* $\{1, 2, \cdots, n\}$ *can be obtained from every other ordered list by a finite number of switches. Also,* sgn *is unique.*

Proof: This is obvious if $n = 1$ or 2. Suppose then that it is true for sets of $n-1$ elements. Take two ordered lists of numbers, P_1, P_2. To get from P_1 to P_2 using switches, first make a switch to obtain the last element in the list coinciding with the last element of P_2. By induction, there are switches which will arrange the first $n-1$ to the right order.

To see sgn_n is unique, if there exist two functions, f and g both satisfying 3.2 and 3.3, you could start with $f(1, \cdots, n) = g(1, \cdots, n)$ and applying the same sequence of switches, eventually arrive at $f(i_1, \cdots, i_n) = g(i_1, \cdots, i_n)$. If any numbers are repeated, then 3.3 gives both functions are equal to zero for that ordered list. ■

Definition 3.3.3 *When you have an ordered list of distinct numbers from* $\{1, 2, \cdots, n\}$, *say* $(i_1, \cdots, i_n)$, *this ordered list is called a permutation. The symbol for all such permutations is* S_n. *The number* $\operatorname{sgn}_n(i_1, \cdots, i_n)$ *is called the sign of the permutation.*

A permutation can also be considered as a function from the set

$$\{1, 2, \cdots, n\} \text{ to } \{1, 2, \cdots, n\}$$

as follows. Let $f(k) = i_k$. Permutations are of fundamental importance in certain areas of math. For example, it was by considering permutations that Galois was able to give a criterion for solution of polynomial equations by radicals, but this is a different direction than what is being attempted here.

In what follows sgn will often be used rather than sgn_n because the context supplies the appropriate n.

3.3.2 The Definition Of The Determinant

Definition 3.3.4 *Let f be a real valued function which has the set of ordered lists of numbers from $\{1, \cdots, n\}$ as its domain. Define*

$$\sum_{(k_1, \cdots, k_n)} f(k_1 \cdots k_n)$$

to be the sum of all the $f(k_1 \cdots k_n)$ for all possible choices of ordered lists $(k_1, \cdots, k_n)$ of numbers of $\{1, \cdots, n\}$. For example,

$$\sum_{(k_1, k_2)} f(k_1, k_2) = f(1,2) + f(2,1) + f(1,1) + f(2,2).$$

Definition 3.3.5 *Let $(a_{ij}) = A$ denote an $n \times n$ matrix. The determinant of A, denoted by $\det(A)$ is defined by*

$$\det(A) \equiv \sum_{(k_1, \cdots, k_n)} \operatorname{sgn}(k_1, \cdots, k_n)\, a_{1k_1} \cdots a_{nk_n}$$

where the sum is taken over all ordered lists of numbers from $\{1, \cdots, n\}$. Note it suffices to take the sum over only those ordered lists in which there are no repeats because if there are, $\operatorname{sgn}(k_1, \cdots, k_n) = 0$ and so that term contributes 0 to the sum.

Let A be an $n \times n$ matrix $A = (a_{ij})$ and let $(r_1, \cdots, r_n)$ denote an ordered list of n numbers from $\{1, \cdots, n\}$. Let $A(r_1, \cdots, r_n)$ denote the matrix whose k^{th} row is the r_k row of the matrix A. Thus

$$\det(A(r_1, \cdots, r_n)) = \sum_{(k_1, \cdots, k_n)} \operatorname{sgn}(k_1, \cdots, k_n)\, a_{r_1 k_1} \cdots a_{r_n k_n} \tag{3.7}$$

and $A(1,\cdots,n) = A$.

Proposition 3.3.6 *Let* $(r_1,\cdots,r_n)$ *be an ordered list of numbers from* $\{1,\cdots,n\}$. *Then*

$$\begin{aligned} \operatorname{sgn}(r_1,\cdots,r_n)\det(A) &= \sum_{(k_1,\cdots,k_n)} \operatorname{sgn}(k_1,\cdots,k_n)\, a_{r_1k_1}\cdots a_{r_nk_n} & (3.8)\\ &= \det(A(r_1,\cdots,r_n)). & (3.9)\end{aligned}$$

Proof: Let $(1,\cdots,n) = (1,\cdots,r,\cdots s,\cdots,n)$ so $r < s$.

$$\det(A(1,\cdots,r,\cdots,s,\cdots,n)) = \tag{3.10}$$

$$\sum_{(k_1,\cdots,k_n)} \operatorname{sgn}(k_1,\cdots,k_r,\cdots,k_s,\cdots,k_n)\, a_{1k_1}\cdots a_{rk_r}\cdots a_{sk_s}\cdots a_{nk_n},$$

and renaming the variables, calling k_s, k_r and k_r, k_s, this equals

$$= \sum_{(k_1,\cdots,k_n)} \operatorname{sgn}(k_1,\cdots,k_s,\cdots,k_r,\cdots,k_n)\, a_{1k_1}\cdots a_{rk_s}\cdots a_{sk_r}\cdots a_{nk_n}$$

$$= \sum_{(k_1,\cdots,k_n)} -\operatorname{sgn}\left(k_1,\cdots,\overbrace{k_r,\cdots,k_s}^{\text{These got switched}},\cdots,k_n\right) a_{1k_1}\cdots a_{sk_r}\cdots a_{rk_s}\cdots a_{nk_n}$$
$$= -\det\left(A\left(1,\cdots,s,\cdots,r,\cdots,n\right)\right). \tag{3.11}$$

Consequently,

$$\det\left(A\left(1,\cdots,s,\cdots,r,\cdots,n\right)\right) = -\det\left(A\left(1,\cdots,r,\cdots,s,\cdots,n\right)\right) = -\det\left(A\right)$$

Now letting $A\left(1,\cdots,s,\cdots,r,\cdots,n\right)$ play the role of A, and continuing in this way, switching pairs of numbers,

$$\det\left(A\left(r_1,\cdots,r_n\right)\right) = (-1)^p \det\left(A\right)$$

where it took p switches to obtain$(r_1,\cdots,r_n)$ from $(1,\cdots,n)$. By Lemma 3.3.1, this implies

$$\det\left(A\left(r_1,\cdots,r_n\right)\right) = (-1)^p \det\left(A\right) = \operatorname{sgn}\left(r_1,\cdots,r_n\right)\det\left(A\right)$$

and proves the proposition in the case when there are no repeated numbers in the ordered list, $(r_1,\cdots,r_n)$. However, if there is a repeat, say the r^{th} row equals the s^{th} row, then the reasoning of 3.10 -3.11 shows that $\det(A\left(r_1,\cdots,r_n\right)) = 0$ and also $\operatorname{sgn}\left(r_1,\cdots,r_n\right) = 0$ so the formula holds in this case also. ■

Observation 3.3.7 *There are $n!$ ordered lists of distinct numbers from $\{1,\cdots,n\}$.*

To see this, consider n slots placed in order. There are n choices for the first slot. For each of these choices, there are $n-1$ choices for the second. Thus there are $n\left(n-1\right)$ ways to fill the first two slots. Then for each of these ways there are $n-2$ choices left for the third slot. Continuing this way, there are $n!$ ordered lists of distinct numbers from $\{1,\cdots,n\}$ as stated in the observation.

3.3.3 A Symmetric Definition

With the above, it is possible to give a more symmetric description of the determinant from which it will follow that $\det\left(A\right) = \det\left(A^T\right)$.

Corollary 3.3.8 *The following formula for $\det\left(A\right)$ is valid.*

$$\det\left(A\right) = \frac{1}{n!}\cdot\sum_{(r_1,\cdots,r_n)}\sum_{(k_1,\cdots,k_n)} \operatorname{sgn}\left(r_1,\cdots,r_n\right)\operatorname{sgn}\left(k_1,\cdots,k_n\right) a_{r_1k_1}\cdots a_{r_nk_n}. \tag{3.12}$$

And also $\det\left(A^T\right) = \det\left(A\right)$ where A^T is the transpose of A. (Recall that for $A^T = \left(a_{ij}^T\right)$, $a_{ij}^T = a_{ji}$.)

Proof: From Proposition 3.3.6, if the r_i are distinct,

$$\det\left(A\right) = \sum_{(k_1,\cdots,k_n)} \operatorname{sgn}\left(r_1,\cdots,r_n\right)\operatorname{sgn}\left(k_1,\cdots,k_n\right) a_{r_1k_1}\cdots a_{r_nk_n}.$$

Summing over all ordered lists, $(r_1,\cdots,r_n)$ where the r_i are distinct, (If the r_i are not distinct, $\operatorname{sgn}\left(r_1,\cdots,r_n\right) = 0$ and so there is no contribution to the sum.)

$$n!\det\left(A\right) = \sum_{(r_1,\cdots,r_n)}\sum_{(k_1,\cdots,k_n)} \operatorname{sgn}\left(r_1,\cdots,r_n\right)\operatorname{sgn}\left(k_1,\cdots,k_n\right) a_{r_1k_1}\cdots a_{r_nk_n}.$$

This proves the corollary since the formula gives the same number for A as it does for A^T. ■

Corollary 3.3.9 *If two rows or two columns in an $n \times n$ matrix A, are switched, the determinant of the resulting matrix equals (-1) times the determinant of the original matrix. If A is an $n \times n$ matrix in which two rows are equal or two columns are equal then* $\det(A) = 0$. *Suppose the i^{th} row of A equals $(xa_1 + yb_1, \cdots, xa_n + yb_n)$. Then*

$$\det(A) = x \det(A_1) + y \det(A_2)$$

where the i^{th} row of A_1 is $(a_1, \cdots, a_n)$ and the i^{th} row of A_2 is $(b_1, \cdots, b_n)$, all other rows of A_1 and A_2 coinciding with those of A. In other words, det *is a linear function of each row A. The same is true with the word "row" replaced with the word "column".*

Proof: By Proposition 3.3.6 when two rows are switched, the determinant of the resulting matrix is (-1) times the determinant of the original matrix. By Corollary 3.3.8 the same holds for columns because the columns of the matrix equal the rows of the transposed matrix. Thus if A_1 is the matrix obtained from A by switching two columns,

$$\det(A) = \det(A^T) = -\det(A_1^T) = -\det(A_1).$$

If A has two equal columns or two equal rows, then switching them results in the same matrix. Therefore, $\det(A) = -\det(A)$ and so $\det(A) = 0$.

It remains to verify the last assertion.

$$\det(A) \equiv \sum_{(k_1,\cdots,k_n)} \operatorname{sgn}(k_1, \cdots, k_n) a_{1k_1} \cdots (xa_{rk_i} + yb_{rk_i}) \cdots a_{nk_n}$$

$$= x \sum_{(k_1,\cdots,k_n)} \operatorname{sgn}(k_1, \cdots, k_n) a_{1k_1} \cdots a_{rk_i} \cdots a_{nk_n}$$

$$+y \sum_{(k_1,\cdots,k_n)} \operatorname{sgn}(k_1, \cdots, k_n) a_{1k_1} \cdots b_{rk_i} \cdots a_{nk_n} \equiv x \det(A_1) + y \det(A_2).$$

The same is true of columns because $\det(A^T) = \det(A)$ and the rows of A^T are the columns of A. ■

3.3.4 Basic Properties Of The Determinant

Definition 3.3.10 *A vector, $\mathbf{w}$, is a linear combination of the vectors $\{\mathbf{v}_1, \cdots, \mathbf{v}_r\}$ if there exist scalars $c_1, \cdots c_r$ such that $\mathbf{w} = \sum_{k=1}^{r} c_k \mathbf{v}_k$. This is the same as saying*

$$\mathbf{w} \in \operatorname{span}(\mathbf{v}_1, \cdots, \mathbf{v}_r).$$

The following corollary is also of great use.

Corollary 3.3.11 *Suppose A is an $n \times n$ matrix and some column (row) is a linear combination of r other columns (rows). Then* $\det(A) = 0$.

Proof: Let $A = \begin{pmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{pmatrix}$ be the columns of A and suppose the condition that one column is a linear combination of r of the others is satisfied. Then by using Corollary 3.3.9 you may rearrange the columns to have the n^{th} column a linear combination of the first r columns. Thus $\mathbf{a}_n = \sum_{k=1}^{r} c_k \mathbf{a}_k$ and so

$$\det(A) = \det \begin{pmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_r & \cdots & \mathbf{a}_{n-1} & \sum_{k=1}^{r} c_k \mathbf{a}_k \end{pmatrix}.$$

By Corollary 3.3.9

$$\det(A) = \sum_{k=1}^{r} c_k \det\begin{pmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_r & \cdots & \mathbf{a}_{n-1} & \mathbf{a}_k \end{pmatrix} = 0.$$

The case for rows follows from the fact that $\det(A) = \det\left(A^T\right)$. ■

Recall the following definition of matrix multiplication.

Definition 3.3.12 *If A and B are $n \times n$ matrices, $A = (a_{ij})$ and $B = (b_{ij})$, $AB = (c_{ij})$ where $c_{ij} \equiv \sum_{k=1}^{n} a_{ik} b_{kj}$.*

One of the most important rules about determinants is that the determinant of a product equals the product of the determinants.

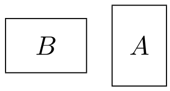
Theorem 3.3.13 *Let A and B be $n \times n$ matrices. Then*

$$\det(AB) = \det(A)\det(B).$$

Proof: Let c_{ij} be the ij^{th} entry of AB. Then by Proposition 3.3.6,

$$\begin{aligned}
\det(AB) &= \sum_{(k_1,\cdots,k_n)} \operatorname{sgn}(k_1,\cdots,k_n)\, c_{1k_1}\cdots c_{nk_n} \\
&= \sum_{(k_1,\cdots,k_n)} \operatorname{sgn}(k_1,\cdots,k_n) \left(\sum_{r_1} a_{1r_1} b_{r_1k_1}\right)\cdots\left(\sum_{r_n} a_{nr_n} b_{r_nk_n}\right) \\
&= \sum_{(r_1\cdots,r_n)} \sum_{(k_1,\cdots,k_n)} \operatorname{sgn}(k_1,\cdots,k_n)\, b_{r_1k_1}\cdots b_{r_nk_n} \left(a_{1r_1}\cdots a_{nr_n}\right) \\
&= \sum_{(r_1\cdots,r_n)} \operatorname{sgn}(r_1\cdots r_n)\, a_{1r_1}\cdots a_{nr_n} \det(B) = \det(A)\det(B). ■
\end{aligned}$$

The Binet Cauchy formula is a generalization of the theorem which says the determinant of a product is the product of the determinants. The situation is illustrated in the following picture where A, B are matrices.



Theorem 3.3.14 *Let A be an $n \times m$ matrix with $n \geq m$ and let B be a $m \times n$ matrix. Also let A_i*

$$i = 1, \cdots, C(n, m)$$

be the $m \times m$ submatrices of A which are obtained by deleting $n - m$ rows and let B_i be the $m \times m$ submatrices of B which are obtained by deleting corresponding $n - m$ columns. Then

$$\det(BA) = \sum_{k=1}^{C(n,m)} \det(B_k) \det(A_k)$$

Proof: This follows from a computation. By Corollary 3.3.8 on Page 113, $\det(BA) =$

$$\frac{1}{m!}\sum_{(i_1\cdots i_m)}\sum_{(j_1\cdots j_m)} \operatorname{sgn}(i_1\cdots i_m)\operatorname{sgn}(j_1\cdots j_m)(BA)_{i_1j_1}(BA)_{i_2j_2}\cdots(BA)_{i_mj_m}$$

$$\frac{1}{m!}\sum_{(i_1\cdots i_m)}\sum_{(j_1\cdots j_m)} \operatorname{sgn}(i_1\cdots i_m)\operatorname{sgn}(j_1\cdots j_m)\cdot$$
$$\sum_{r_1=1}^{n} B_{i_1r_1}A_{r_1j_1}\sum_{r_2=1}^{n} B_{i_2r_2}A_{r_2j_2}\cdots\sum_{r_m=1}^{n} B_{i_mr_m}A_{r_mj_m}$$

Now denote by I_k one of the r subsets of $\{1,\cdots,n\}$. Thus there are $C(n,m)$ of these.

$$=\sum_{k=1}^{C(n,m)}\sum_{\{r_1,\cdots,r_m\}=I_k}\frac{1}{m!}\sum_{(i_1\cdots i_m)}\sum_{(j_1\cdots j_m)} \operatorname{sgn}(i_1\cdots i_m)\operatorname{sgn}(j_1\cdots j_m)\cdot$$
$$B_{i_1r_1}A_{r_1j_1}B_{i_2r_2}A_{r_2j_2}\cdots B_{i_mr_m}A_{r_mj_m}$$

$$=\sum_{k=1}^{C(n,m)}\sum_{\{r_1,\cdots,r_m\}=I_k}\frac{1}{m!}\sum_{(i_1\cdots i_m)} \operatorname{sgn}(i_1\cdots i_m)B_{i_1r_1}B_{i_2r_2}\cdots B_{i_mr_m}\cdot$$
$$\sum_{(j_1\cdots j_m)}\operatorname{sgn}(j_1\cdots j_m)A_{r_1j_1}A_{r_2j_2}\cdots A_{r_mj_m}$$

$$=\sum_{k=1}^{C(n,m)}\sum_{\{r_1,\cdots,r_m\}=I_k}\frac{1}{m!}\operatorname{sgn}(r_1\cdots r_m)^2\det(B_k)\det(A_k)\,B$$
$$=\sum_{k=1}^{C(n,m)}\det(B_k)\det(A_k)$$

since there are $m!$ ways of arranging the indices $\{r_1,\cdots,r_m\}$. ■

3.3.5 Expansion Using Cofactors

Lemma 3.3.15 *Suppose a matrix is of the form*

$$M=\begin{pmatrix} A & * \\ \mathbf{0} & a\end{pmatrix} \tag{3.13}$$

or

$$M=\begin{pmatrix} A & \mathbf{0} \\ * & a\end{pmatrix} \tag{3.14}$$

where a is a number and A is an $(n-1)\times(n-1)$ matrix and $*$ denotes either a column or a row having length $n-1$ and the $\mathbf{0}$ denotes either a column or a row of length $n-1$ consisting entirely of zeros. Then $\det(M) = a\det(A)$.

Proof: Denote M by (m_{ij}). Thus in the first case, $m_{nn} = a$ and $m_{ni} = 0$ if $i \neq n$ while in the second case, $m_{nn} = a$ and $m_{in} = 0$ if $i \neq n$. From the definition of the determinant,

$$\det(M) \equiv \sum_{(k_1,\cdots,k_n)}\operatorname{sgn}_n(k_1,\cdots,k_n)\,m_{1k_1}\cdots m_{nk_n}$$

Letting θ denote the position of n in the ordered list, $(k_1, \cdots, k_n)$ then using the earlier conventions used to prove Lemma 3.3.1, $\det(M)$ equals

$$\sum_{(k_1,\cdots,k_n)} (-1)^{n-\theta} \operatorname{sgn}_{n-1}\left(k_1, \cdots, k_{\theta-1}, \overset{\theta}{k_{\theta+1}}, \cdots, \overset{n-1}{k_n}\right) m_{1k_1} \cdots m_{nk_n}$$

Now suppose 3.14. Then if $k_n \neq n$, the term involving m_{nk_n} in the above expression equals zero. Therefore, the only terms which survive are those for which $\theta = n$ or in other words, those for which $k_n = n$. Therefore, the above expression reduces to

$$a \sum_{(k_1,\cdots,k_{n-1})} \operatorname{sgn}_{n-1}(k_1, \cdots k_{n-1}) m_{1k_1} \cdots m_{(n-1)k_{n-1}} = a \det(A).$$

To get the assertion in the situation of 3.13 use Corollary 3.3.8 and 3.14 to write

$$\det(M) = \det\left(M^T\right) = \det\left(\begin{pmatrix} A^T & \mathbf{0} \\ * & a \end{pmatrix}\right) = a \det\left(A^T\right) = a \det(A). \blacksquare$$

In terms of the theory of determinants, arguably the most important idea is that of Laplace expansion along a row or a column. This will follow from the above definition of a determinant.

Definition 3.3.16 *Let $A = (a_{ij})$ be an $n \times n$ matrix. Then a new matrix called the cofactor matrix $\operatorname{cof}(A)$ is defined by $\operatorname{cof}(A) = (c_{ij})$ where to obtain c_{ij} delete the i^{th} row and the j^{th} column of A, take the determinant of the $(n-1) \times (n-1)$ matrix which results, (This is called the ij^{th} minor of A.) and then multiply this number by $(-1)^{i+j}$. To make the formulas easier to remember, $\operatorname{cof}(A)_{ij}$ will denote the ij^{th} entry of the cofactor matrix.*

The following is the main result. Earlier this was given as a definition and the outrageous totally unjustified assertion was made that the same number would be obtained by expanding the determinant along any row or column. The following theorem proves this assertion.

Theorem 3.3.17 *Let A be an $n \times n$ matrix where $n \geq 2$. Then*

$$\det(A) = \sum_{j=1}^{n} a_{ij} \operatorname{cof}(A)_{ij} = \sum_{i=1}^{n} a_{ij} \operatorname{cof}(A)_{ij}. \tag{3.15}$$

The first formula consists of expanding the determinant along the i^{th} row and the second expands the determinant along the j^{th} column.

Proof: Let $(a_{i1}, \cdots, a_{in})$ be the i^{th} row of A. Let B_j be the matrix obtained from A by leaving every row the same except the i^{th} row which in B_j equals $(0, \cdots, 0, a_{ij}, 0, \cdots, 0)$. Then by Corollary 3.3.9,

$$\det(A) = \sum_{j=1}^{n} \det(B_j)$$

For example if

$$A = \begin{pmatrix} a & b & c \\ d & e & f \\ h & i & j \end{pmatrix}$$

and $i = 2$, then

$$B_1 = \begin{pmatrix} a & b & c \\ d & 0 & 0 \\ h & i & j \end{pmatrix}, B_2 = \begin{pmatrix} a & b & c \\ 0 & e & 0 \\ h & i & j \end{pmatrix}, B_3 = \begin{pmatrix} a & b & c \\ 0 & 0 & f \\ h & i & j \end{pmatrix}$$

Denote by A^{ij} the $(n-1)\times(n-1)$ matrix obtained by deleting the i^{th} row and the j^{th} column of A. Thus $\operatorname{cof}(A)_{ij} \equiv (-1)^{i+j}\det\left(A^{ij}\right)$. At this point, recall that from Proposition 3.3.6, when two rows or two columns in a matrix M, are switched, this results in multiplying the determinant of the old matrix by -1 to get the determinant of the new matrix. Therefore, by Lemma 3.3.15,

$$\begin{aligned}\det(B_j) &= (-1)^{n-j}(-1)^{n-i}\det\left(\begin{pmatrix} A^{ij} & * \\ \mathbf{0} & a_{ij}\end{pmatrix}\right) \\ &= (-1)^{i+j}\det\left(\begin{pmatrix} A^{ij} & * \\ \mathbf{0} & a_{ij}\end{pmatrix}\right) = a_{ij}\operatorname{cof}(A)_{ij}.\end{aligned}$$

Therefore,

$$\det(A) = \sum_{j=1}^{n} a_{ij}\operatorname{cof}(A)_{ij}$$

which is the formula for expanding $\det(A)$ along the i^{th} row. Also,

$$\det(A) = \det\left(A^T\right) = \sum_{j=1}^{n} a_{ij}^T\operatorname{cof}\left(A^T\right)_{ij} = \sum_{j=1}^{n} a_{ji}\operatorname{cof}(A)_{ji}$$

which is the formula for expanding $\det(A)$ along the i^{th} column. ■

3.3.6 A Formula For The Inverse

Note that this gives an easy way to write a formula for the inverse of an $n\times n$ matrix. Recall the definition of the inverse of a matrix in Definition 2.1.22 on Page 61.

Theorem 3.3.18 *A^{-1} exists if and only if $\det(A) \neq 0$. If $\det(A) \neq 0$, then $A^{-1} = \left(a_{ij}^{-1}\right)$ where*

$$a_{ij}^{-1} = \det(A)^{-1}\operatorname{cof}(A)_{ji}$$

for $\operatorname{cof}(A)_{ij}$ the ij^{th} cofactor of A.

Proof: By Theorem 3.3.17 and letting $(a_{ir}) = A$, if $\det(A) \neq 0$,

$$\sum_{i=1}^{n} a_{ir}\operatorname{cof}(A)_{ir}\det(A)^{-1} = \det(A)\det(A)^{-1} = 1.$$

Now in the matrix A, replace the k^{th} column with the r^{th} column and then expand along the k^{th} column. This yields for $k \neq r$,

$$\sum_{i=1}^{n} a_{ir}\operatorname{cof}(A)_{ik}\det(A)^{-1} = 0$$

because there are two equal columns by Corollary 3.3.9. Summarizing,

$$\sum_{i=1}^{n} a_{ir}\operatorname{cof}(A)_{ik}\det(A)^{-1} = \delta_{rk}.$$

Using the other formula in Theorem 3.3.17, and similar reasoning,

$$\sum_{j=1}^{n} a_{rj}\operatorname{cof}(A)_{kj}\det(A)^{-1} = \delta_{rk}$$

This proves that if $\det(A) \neq 0$, then A^{-1} exists with $A^{-1} = \left(a_{ij}^{-1}\right)$, where

$$a_{ij}^{-1} = \operatorname{cof}(A)_{ji} \det(A)^{-1}.$$

Now suppose A^{-1} exists. Then by Theorem 3.3.13,

$$1 = \det(I) = \det\left(AA^{-1}\right) = \det(A)\det\left(A^{-1}\right)$$

so $\det(A) \neq 0$. ■

The next corollary points out that if an $n \times n$ matrix A has a right or a left inverse, then it has an inverse.

Corollary 3.3.19 *Let A be an $n \times n$ matrix and suppose there exists an $n \times n$ matrix B such that $BA = I$. Then A^{-1} exists and $A^{-1} = B$. Also, if there exists C an $n \times n$ matrix such that $AC = I$, then A^{-1} exists and $A^{-1} = C$.*

Proof: Since $BA = I$, Theorem 3.3.13 implies

$$\det B \det A = 1$$

and so $\det A \neq 0$. Therefore from Theorem 3.3.18, A^{-1} exists. Therefore,

$$A^{-1} = (BA)A^{-1} = B\left(AA^{-1}\right) = BI = B.$$

The case where $CA = I$ is handled similarly. ■

The conclusion of this corollary is that left inverses, right inverses and inverses are all the same in the context of $n \times n$ matrices.

Theorem 3.3.18 says that to find the inverse, take the transpose of the cofactor matrix and divide by the determinant. The transpose of the cofactor matrix is called the adjugate or sometimes the classical adjoint of the matrix A. It is an abomination to call it the adjoint although you do sometimes see it referred to in this way. In words, A^{-1} is equal to one over the determinant of A times the adjugate matrix of A.

In case you are solving a system of equations, $A\mathbf{x} = \mathbf{y}$ for $\mathbf{x}$, it follows that if A^{-1} exists,

$$\mathbf{x} = \left(A^{-1}A\right)\mathbf{x} = A^{-1}\left(A\mathbf{x}\right) = A^{-1}\mathbf{y}$$

thus solving the system. Now in the case that A^{-1} exists, there is a formula for A^{-1} given above. Using this formula,

$$x_i = \sum_{j=1}^{n} a_{ij}^{-1} y_j = \sum_{j=1}^{n} \frac{1}{\det\left(A\right)} \operatorname{cof}\left(A\right)_{ji} y_j.$$

By the formula for the expansion of a determinant along a column,

$$x_i = \frac{1}{\det(A)} \det \begin{pmatrix} * & \cdots & y_1 & \cdots & * \\ \vdots & & \vdots & & \vdots \\ * & \cdots & y_n & \cdots & * \end{pmatrix},$$

where here the i^{th} column of A is replaced with the column vector, $(y_1 \cdots, y_n)^T$, and the determinant of this modified matrix is taken and divided by $\det(A)$. This formula is known as Cramer's rule.

Definition 3.3.20 *A matrix M, is upper triangular if $M_{ij} = 0$ whenever $i > j$. Thus such a matrix equals zero below the main diagonal, the entries of the form M_{ii} as shown.*

$$\begin{pmatrix} * & * & \cdots & * \\ 0 & * & \ddots & \vdots \\ \vdots & \ddots & \ddots & * \\ 0 & \cdots & 0 & * \end{pmatrix}$$

A lower triangular matrix is defined similarly as a matrix for which all entries above the main diagonal are equal to zero.

With this definition, here is a simple corollary of Theorem 3.3.17.

Corollary 3.3.21 *Let M be an upper (lower) triangular matrix. Then $\det(M)$ is obtained by taking the product of the entries on the main diagonal.*

3.3.7 Rank Of A Matrix

Definition 3.3.22 *A submatrix of a matrix A is the rectangular array of numbers obtained by deleting some rows and columns of A. Let A be an $m \times n$ matrix. The* ***determinant rank*** *of the matrix equals r where r is the largest number such that some $r \times r$ submatrix of A has a non zero determinant. The* ***row rank*** *is defined to be the dimension of the span of the rows. The* ***column rank*** *is defined to be the dimension of the span of the columns.*

Theorem 3.3.23 *If A, an $m \times n$ matrix has determinant rank r, then there exist r rows of the matrix such that every other row is a linear combination of these r rows.*

Proof: Suppose the determinant rank of $A = (a_{ij})$ equals r. Thus some $r \times r$ submatrix has non zero determinant and there is no larger square submatrix which has non zero determinant. Suppose such a submatrix is determined by the r columns whose indices are

$$j_1 < \cdots < j_r$$

and the r rows whose indices are

$$i_1 < \cdots < i_r$$

I want to show that every row is a linear combination of these rows. Consider the l^{th} row and let p be an index between 1 and n. Form the following $(r+1) \times (r+1)$ matrix

$$\begin{pmatrix} a_{i_1 j_1} & \cdots & a_{i_1 j_r} & a_{i_1 p} \\ \vdots & & \vdots & \vdots \\ a_{i_r j_1} & \cdots & a_{i_r j_r} & a_{i_r p} \\ a_{l j_1} & \cdots & a_{l j_r} & a_{l p} \end{pmatrix}$$

Of course you can assume $l \notin \{i_1, \cdots, i_r\}$ because there is nothing to prove if the l^{th} row is one of the chosen ones. The above matrix has determinant 0. This is because if $p \notin \{j_1, \cdots, j_r\}$ then the above would be a submatrix of A which is too large to have non zero determinant. On the other hand, if $p \in \{j_1, \cdots, j_r\}$ then the above matrix has two columns which are equal so its determinant is still 0.

Expand the determinant of the above matrix along the last column. Let C_k denote the cofactor associated with the entry $a_{i_k p}$. This is not dependent on the choice of p. Remember, you delete the column and the row the entry is in and take the determinant of what is left and multiply by -1 raised to an appropriate power. Let C denote the cofactor associated with a_{lp}. This is given to be nonzero, it being the determinant of the matrix

$$\begin{pmatrix} a_{i_1 j_1} & \cdots & a_{i_1 j_r} \\ \vdots & & \vdots \\ a_{i_r j_1} & \cdots & a_{i_r j_r} \end{pmatrix}$$

Thus

$$0 = a_{lp}C + \sum_{k=1}^{r} C_k a_{i_k p}$$

which implies

$$a_{lp} = \sum_{k=1}^{r} \frac{-C_k}{C} a_{i_k p} \equiv \sum_{k=1}^{r} m_k a_{i_k p}$$

Since this is true for every p and since m_k does not depend on p, this has shown the l^{th} row is a linear combination of the $i_1, i_2, \cdots, i_r$ rows. ■

Corollary 3.3.24 *The determinant rank equals the row rank.*

Proof: From Theorem 3.3.23, every row is in the span of r rows where r is the determinant rank. Therefore, the row rank (dimension of the span of the rows) is no larger than the determinant rank. Could the row rank be smaller than the determinant rank? If so, it follows from Theorem 3.3.23 that there exist p rows for $p < r \equiv$ determinant rank, such that the span of these p rows equals the row space. But then you could consider the $r \times r$ sub matrix which determines the determinant rank and it would follow that each of these rows would be in the span of the restrictions of the p rows just mentioned. By Theorem 2.4.4, the exchange theorem, the rows of this sub matrix would not be linearly independent and so some row is a linear combination of the others. By Corollary 3.3.11 the determinant would be 0, a contradiction. ■

Corollary 3.3.25 *If A has determinant rank r, then there exist r columns of the matrix such that every other column is a linear combination of these r columns. Also the column rank equals the determinant rank.*

Proof: This follows from the above by considering A^T. The rows of A^T are the columns of A and the determinant rank of A^T and A are the same. Therefore, from Corollary 3.3.24, column rank of $A =$ row rank of $A^T =$ determinant rank of $A^T =$ determinant rank of A.

The following theorem is of fundamental importance and ties together many of the ideas presented above.

Theorem 3.3.26 *Let A be an $n \times n$ matrix. Then the following are equivalent.*

1. $\det(A) = 0$.

2. *A, A^T are not one to one.*

3. *A is not onto.*

Proof: Suppose $\det(A) = 0$. Then the determinant rank of $A = r < n$. Therefore, there exist r columns such that every other column is a linear combination of these columns by Theorem 3.3.23. In particular, it follows that for some m, the m^{th} column is a linear combination of all the others. Thus letting $A = \begin{pmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_m & \cdots & \mathbf{a}_n \end{pmatrix}$ where the columns are denoted by $\mathbf{a}_i$, there exists scalars α_i such that

$$\mathbf{a}_m = \sum_{k \neq m} \alpha_k \mathbf{a}_k.$$

Now consider the column vector, $\mathbf{x} \equiv \begin{pmatrix} \alpha_1 & \cdots & -1 & \cdots & \alpha_n \end{pmatrix}^T$. Then

$$A\mathbf{x} = -\mathbf{a}_m + \sum_{k \neq m} \alpha_k \mathbf{a}_k = \mathbf{0}.$$

Since also $A\mathbf{0} = \mathbf{0}$, it follows A is not one to one. Similarly, A^T is not one to one by the same argument applied to A^T. This verifies that 1.) implies 2.).

Now suppose 2.). Then since A^T is not one to one, it follows there exists $\mathbf{x} \neq \mathbf{0}$ such that

$$A^T \mathbf{x} = \mathbf{0}.$$

Taking the transpose of both sides yields

$$\mathbf{x}^T A = \mathbf{0}^T$$

where the $\mathbf{0}^T$ is a $1 \times n$ matrix or row vector. Now if $A\mathbf{y} = \mathbf{x}$, then

$$|\mathbf{x}|^2 = \mathbf{x}^T (A\mathbf{y}) = \left(\mathbf{x}^T A\right) \mathbf{y} = \mathbf{0}\mathbf{y} = 0$$

contrary to $\mathbf{x} \neq \mathbf{0}$. Consequently there can be no $\mathbf{y}$ such that $A\mathbf{y} = \mathbf{x}$ and so A is not onto. This shows that 2.) implies 3.).

Finally, suppose 3.). If 1.) does not hold, then $\det(A) \neq 0$ but then from Theorem 3.3.18 A^{-1} exists and so for every $\mathbf{y} \in \mathbb{F}^n$ there exists a unique $\mathbf{x} \in \mathbb{F}^n$ such that $A\mathbf{x} = \mathbf{y}$. In fact

$\mathbf{x} = A^{-1}\mathbf{y}$. Thus A would be onto contrary to 3.). This shows 3.) implies 1.). ■

Corollary 3.3.27 *Let A be an $n \times n$ matrix. Then the following are equivalent.*

1. $det(A) \neq 0$.
2. *A and A^T are one to one.*
3. *A is onto.*

Proof: This follows immediately from the above theorem.

3.3.8 Summary Of Determinants

In all the following A, B are $n \times n$ matrices

1. $\det(A)$ is a number.
2. $\det(A)$ is linear in each row and in each column.

3. If you switch two rows or two columns, the determinant of the resulting matrix is -1 times the determinant of the unswitched matrix. (This and the previous one say

$$(\mathbf{a}_1 \cdots \mathbf{a}_n) \rightarrow \det(\mathbf{a}_1 \cdots \mathbf{a}_n)$$

is an alternating multilinear function or alternating tensor.

4. $\det(\mathbf{e}_1, \cdots, \mathbf{e}_n) = 1$.

5. $\det(AB) = \det(A)\det(B)$

6. $\det(A)$ can be expanded along any row or any column and the same result is obtained.

7. $\det(A) = \det\left(A^T\right)$

8. A^{-1} exists if and only if $\det(A) \neq 0$ and in this case

$$\left(A^{-1}\right)_{ij} = \frac{1}{\det(A)} \operatorname{cof}(A)_{ji} \tag{3.16}$$

9. Determinant rank, row rank and column rank are all the same number for any $m \times n$ matrix.

3.4 The Cayley Hamilton Theorem

Definition 3.4.1 *Let A be an $n \times n$ matrix. The characteristic polynomial is defined as*

$$p_A(t) \equiv \det(tI - A)$$

and the solutions to $p_A(t) = 0$ are called eigenvalues. For A a matrix and $p(t) = t^n + a_{n-1}t^{n-1} + \cdots + a_1 t + a_0$, denote by $p(A)$ the matrix defined by

$$p(A) \equiv A^n + a_{n-1}A^{n-1} + \cdots + a_1 A + a_0 I.$$

The explanation for the last term is that A^0 is interpreted as I, the identity matrix.

The Cayley Hamilton theorem states that every matrix satisfies its characteristic equation, that equation defined by $p_A(t) = 0$. It is one of the most important theorems in linear algebra[1]. The following lemma will help with its proof.

Lemma 3.4.2 *Suppose for all $|\lambda|$ large enough,*

$$A_0 + A_1\lambda + \cdots + A_m\lambda^m = 0,$$

where the A_i are $n \times n$ matrices. Then each $A_i = 0$.

Proof: Multiply by λ^{-m} to obtain

$$A_0\lambda^{-m} + A_1\lambda^{-m+1} + \cdots + A_{m-1}\lambda^{-1} + A_m = 0.$$

Now let $|\lambda| \rightarrow \infty$ to obtain $A_m = 0$. With this, multiply by λ to obtain

$$A_0\lambda^{-m+1} + A_1\lambda^{-m+2} + \cdots + A_{m-1} = 0.$$

Now let $|\lambda| \rightarrow \infty$ to obtain $A_{m-1} = 0$. Continue multiplying by λ and letting $\lambda \rightarrow \infty$ to obtain that all the $A_i = 0$. ■

With the lemma, here is a simple corollary.

[1] A special case was first proved by Hamilton in 1853. The general case was announced by Cayley some time later and a proof was given by Frobenius in 1878.

Corollary 3.4.3 *Let A_i and B_i be $n \times n$ matrices and suppose*

$$A_0 + A_1\lambda + \cdots + A_m\lambda^m = B_0 + B_1\lambda + \cdots + B_m\lambda^m$$

for all $|\lambda|$ large enough. Then $A_i = B_i$ for all i. Consequently if λ is replaced by any $n \times n$ matrix, the two sides will be equal. That is, for C any $n \times n$ matrix,

$$A_0 + A_1C + \cdots + A_mC^m = B_0 + B_1C + \cdots + B_mC^m.$$

Proof: Subtract and use the result of the lemma. ■

With this preparation, here is a relatively easy proof of the Cayley Hamilton theorem.

Theorem 3.4.4 *Let A be an $n \times n$ matrix and let $p(\lambda) \equiv \det(\lambda I - A)$ be the characteristic polynomial. Then $p(A) = 0$.*

Proof: Let $C(\lambda)$ equal the transpose of the cofactor matrix of $(\lambda I - A)$ for $|\lambda|$ large. (If $|\lambda|$ is large enough, then λ cannot be in the finite list of eigenvalues of A and so for such λ, $(\lambda I - A)^{-1}$ exists.) Therefore, by Theorem 3.3.18

$$C(\lambda) = p(\lambda)(\lambda I - A)^{-1}.$$

Note that each entry in $C(\lambda)$ is a polynomial in λ having degree no more than $n-1$. Therefore, collecting the terms,

$$C(\lambda) = C_0 + C_1\lambda + \cdots + C_{n-1}\lambda^{n-1}$$

for C_j some $n \times n$ matrix. It follows that for all $|\lambda|$ large enough,

$$(\lambda I - A)\left(C_0 + C_1\lambda + \cdots + C_{n-1}\lambda^{n-1}\right) = p(\lambda) I$$

and so Corollary 3.4.3 may be used. It follows the matrix coefficients corresponding to equal powers of λ are equal on both sides of this equation. Therefore, if λ is replaced with A, the two sides will be equal. Thus

$$0 = (A - A)\left(C_0 + C_1A + \cdots + C_{n-1}A^{n-1}\right) = p(A) I = p(A). ■$$

3.5 Block Multiplication Of Matrices

Consider the following problem

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}\begin{pmatrix} E & F \\ G & H \end{pmatrix}$$

You know how to do this. You get

$$\begin{pmatrix} AE + BG & AF + BH \\ CE + DG & CF + DH \end{pmatrix}.$$

Now what if instead of numbers, the entries, A, B, C, D, E, F, G are matrices of a size such that the multiplications and additions needed in the above formula all make sense. Would the formula be true in this case? I will show below that this is true.

Suppose A is a matrix of the form

$$A = \begin{pmatrix} A_{11} & \cdots & A_{1m} \\ \vdots & \ddots & \vdots \\ A_{r1} & \cdots & A_{rm} \end{pmatrix} \tag{3.17}$$

where A_{ij} is a $s_i \times p_j$ matrix where s_i is constant for $j = 1, \cdots, m$ for each $i = 1, \cdots, r$. Such a matrix is called a **block matrix,** also a **partitioned matrix**. How do you get the block A_{ij}? Here is how for A an $m \times n$ matrix:

$$\overbrace{\begin{pmatrix} \mathbf{0} & I_{s_i \times s_i} & \mathbf{0} \end{pmatrix}}^{s_i \times m} A \overbrace{\begin{pmatrix} \mathbf{0} \\ I_{p_j \times p_j} \\ \mathbf{0} \end{pmatrix}}^{n \times p_j}. \tag{3.18}$$

In the block column matrix on the right, you need to have $c_j - 1$ rows of zeros above the small $p_j \times p_j$ identity matrix where the columns of A involved in A_{ij} are $c_j, \cdots, c_j + p_j - 1$ and in the block row matrix on the left, you need to have $r_i - 1$ columns of zeros to the left of the $s_i \times s_i$ identity matrix where the rows of A involved in A_{ij} are $r_i, \cdots, r_i + s_i$. An important observation to make is that the matrix on the right specifies columns to use in the block and the one on the left specifies the rows used. Thus the block A_{ij} in this case is a matrix of size $s_i \times p_j$. There is no overlap between the blocks of A. Thus the identity $n \times n$ identity matrix corresponding to multiplication on the right of A is of the form

$$\begin{pmatrix} I_{p_1 \times p_1} & & 0 \\ & \ddots & \\ 0 & & I_{p_m \times p_m} \end{pmatrix}$$

where these little identity matrices don't overlap. A similar conclusion follows from consideration of the matrices $I_{s_i \times s_i}$. Note that in 3.18 the matrix on the right is a block column matrix for the above block diagonal matrix and the matrix on the left in 3.18 is a block row matrix taken from a similar block diagonal matrix consisting of the $I_{s_i \times s_i}$.

Next consider the question of multiplication of two block matrices. Let B be a block matrix of the form

$$\begin{pmatrix} B_{11} & \cdots & B_{1p} \\ \vdots & \ddots & \vdots \\ B_{r1} & \cdots & B_{rp} \end{pmatrix} \tag{3.19}$$

and A is a block matrix of the form

$$\begin{pmatrix} A_{11} & \cdots & A_{1m} \\ \vdots & \ddots & \vdots \\ A_{p1} & \cdots & A_{pm} \end{pmatrix} \tag{3.20}$$

and that for all i, j, it makes sense to multiply $B_{is}A_{sj}$ for all $s \in \{1, \cdots, p\}$. (That is the two matrices, B_{is} and A_{sj} are conformable.) and that for fixed ij, it follows $B_{is}A_{sj}$ is the same size for each s so that it makes sense to write $\sum_s B_{is} A_{sj}$.

The following theorem says essentially that when you take the product of two matrices, you can do it two ways. One way is to simply multiply them forming BA. The other way is to partition both matrices, formally multiply the blocks to get another block matrix and this one will be BA partitioned. Before presenting this theorem, here is a simple lemma which is really a special case of the theorem.

Lemma 3.5.1 *Consider the following product.*

$$\begin{pmatrix} \mathbf{0} \\ I \\ \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{0} & I & \mathbf{0} \end{pmatrix}$$

where the first is $n \times r$ and the second is $r \times n$. The small identity matrix I is an $r \times r$ matrix and there are l zero rows above I and l zero columns to the left of I in the right matrix.

Then the product of these matrices is a block matrix of the form

$$\begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & I & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}$$

Proof: From the definition of the way you multiply matrices, the product is

$$\left(\begin{pmatrix} \mathbf{0} \\ I \\ \mathbf{0} \end{pmatrix} \mathbf{0} \cdots \begin{pmatrix} \mathbf{0} \\ I \\ \mathbf{0} \end{pmatrix} \mathbf{0} \ \begin{pmatrix} \mathbf{0} \\ I \\ \mathbf{0} \end{pmatrix} \mathbf{e}_1 \ \cdots \ \begin{pmatrix} \mathbf{0} \\ I \\ \mathbf{0} \end{pmatrix} \mathbf{e}_r \ \begin{pmatrix} \mathbf{0} \\ I \\ \mathbf{0} \end{pmatrix} \mathbf{0} \ \cdots \ \begin{pmatrix} \mathbf{0} \\ I \\ \mathbf{0} \end{pmatrix} \mathbf{0} \right)$$

which yields the claimed result. In the formula $\mathbf{e}_j$ refers to the column vector of length r which has a 1 in the j^{th} position. ■

Theorem 3.5.2 *Let B be a $q \times p$ block matrix as in 3.19 and let A be a $p \times n$ block matrix as in 3.20 such that B_{is} is conformable with A_{sj} and each product, $B_{is}A_{sj}$ for $s = 1, \cdots, p$ is of the same size so they can be added. Then BA can be obtained as a block matrix such that the ij^{th} block is of the form*

$$\sum_s B_{is}A_{sj}. \tag{3.21}$$

Proof: From 3.18

$$B_{is}A_{sj} = \begin{pmatrix} \mathbf{0} & I_{r_i \times r_i} & \mathbf{0} \end{pmatrix} B \begin{pmatrix} \mathbf{0} \\ I_{p_s \times p_s} \\ \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{0} & I_{p_s \times p_s} & \mathbf{0} \end{pmatrix} A \begin{pmatrix} \mathbf{0} \\ I_{q_j \times q_j} \\ \mathbf{0} \end{pmatrix}$$

where here it is assumed B_{is} is $r_i \times p_s$ and A_{sj} is $p_s \times q_j$. The product involves the s^{th} block in the i^{th} row of blocks for B and the s^{th} block in the j^{th} column of A. Thus there are the same number of rows above the $I_{p_s \times p_s}$ as there are columns to the left of $I_{p_s \times p_s}$ in those two inside matrices. Then from Lemma 3.5.1

$$\begin{pmatrix} \mathbf{0} \\ I_{p_s \times p_s} \\ \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{0} & I_{p_s \times p_s} & \mathbf{0} \end{pmatrix} = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & I_{p_s \times p_s} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}$$

Since the blocks of small identity matrices do not overlap,

$$\sum_s \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & I_{p_s \times p_s} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix} = \begin{pmatrix} I_{p_1 \times p_1} & & 0 \\ & \ddots & \\ 0 & & I_{p_p \times p_p} \end{pmatrix} = I$$

and so

$$\begin{aligned} \sum_s B_{is}A_{sj} &= \sum_s \begin{pmatrix} \mathbf{0} & I_{r_i \times r_i} & \mathbf{0} \end{pmatrix} B \begin{pmatrix} \mathbf{0} \\ I_{p_s \times p_s} \\ \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{0} & I_{p_s \times p_s} & \mathbf{0} \end{pmatrix} A \begin{pmatrix} \mathbf{0} \\ I_{q_j \times q_j} \\ \mathbf{0} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{0} & I_{r_i \times r_i} & \mathbf{0} \end{pmatrix} B \sum_s \begin{pmatrix} \mathbf{0} \\ I_{p_s \times p_s} \\ \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{0} & I_{p_s \times p_s} & \mathbf{0} \end{pmatrix} A \begin{pmatrix} \mathbf{0} \\ I_{q_j \times q_j} \\ \mathbf{0} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{0} & I_{r_i \times r_i} & \mathbf{0} \end{pmatrix} BIA \begin{pmatrix} \mathbf{0} \\ I_{q_j \times q_j} \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} \mathbf{0} & I_{r_i \times r_i} & \mathbf{0} \end{pmatrix} BA \begin{pmatrix} \mathbf{0} \\ I_{q_j \times q_j} \\ \mathbf{0} \end{pmatrix} \end{aligned}$$

which equals the ij^{th} block of BA. Hence the ij^{th} block of BA equals the formal multiplication according to matrix multiplication, $\sum_s B_{is}A_{sj}$. ■

Example 3.5.3 *Let an $n \times n$ matrix have the form $A = \begin{pmatrix} a & \mathbf{b} \\ \mathbf{c} & P \end{pmatrix}$ where P is $n-1 \times n-1$. Multiply it by $B = \begin{pmatrix} p & \mathbf{q} \\ \mathbf{r} & Q \end{pmatrix}$ where B is also an $n \times n$ matrix and Q is $n-1 \times n-1$.*

You use block multiplication

$$\begin{pmatrix} a & \mathbf{b} \\ \mathbf{c} & P \end{pmatrix} \begin{pmatrix} p & \mathbf{q} \\ \mathbf{r} & Q \end{pmatrix} = \begin{pmatrix} ap + \mathbf{br} & a\mathbf{q} + \mathbf{b}Q \\ p\mathbf{c} + P\mathbf{r} & \mathbf{cq} + PQ \end{pmatrix}$$

Note that this all makes sense. For example, $\mathbf{b} = 1 \times n - 1$ and $\mathbf{r} = n - 1 \times 1$ so $\mathbf{br}$ is a 1×1. Similar considerations apply to the other blocks.

Here is an interesting and significant application of block multiplication. In this theorem, $p_M(t)$ denotes the characteristic polynomial, $\det(tI - M)$. The zeros of this polynomial will be shown later to be eigenvalues of the matrix M. First note that from block multiplication, for the following block matrices consisting of square blocks of an appropriate size,

$$\begin{pmatrix} A & 0 \\ B & C \end{pmatrix} = \begin{pmatrix} A & 0 \\ B & I \end{pmatrix} \begin{pmatrix} I & 0 \\ 0 & C \end{pmatrix} \text{ so}$$

$$\det \begin{pmatrix} A & 0 \\ B & C \end{pmatrix} = \det \begin{pmatrix} A & 0 \\ B & I \end{pmatrix} \det \begin{pmatrix} I & 0 \\ 0 & C \end{pmatrix} = \det(A)\det(C)$$

Theorem 3.5.4 *Let A be an $m \times n$ matrix and let B be an $n \times m$ matrix for $m \leq n$. Then*

$$p_{BA}(t) = t^{n-m} p_{AB}(t),$$

so the eigenvalues of BA and AB are the same including multiplicities except that BA has $n-m$ extra zero eigenvalues. Here $p_A(t)$ denotes the characteristic polynomial of the matrix A.

Proof: Use block multiplication to write

$$\begin{pmatrix} AB & 0 \\ B & 0 \end{pmatrix} \begin{pmatrix} I & A \\ 0 & I \end{pmatrix} = \begin{pmatrix} AB & ABA \\ B & BA \end{pmatrix}$$

$$\begin{pmatrix} I & A \\ 0 & I \end{pmatrix} \begin{pmatrix} 0 & 0 \\ B & BA \end{pmatrix} = \begin{pmatrix} AB & ABA \\ B & BA \end{pmatrix}.$$

Therefore,

$$\begin{pmatrix} I & A \\ 0 & I \end{pmatrix}^{-1} \begin{pmatrix} AB & 0 \\ B & 0 \end{pmatrix} \begin{pmatrix} I & A \\ 0 & I \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ B & BA \end{pmatrix}$$

Since the two matrices above are similar, it follows that $\begin{pmatrix} 0 & 0 \\ B & BA \end{pmatrix}$ and $\begin{pmatrix} AB & 0 \\ B & 0 \end{pmatrix}$ have the same characteristic polynomials. See Problem 8 on Page 106. Therefore, noting that BA is an $n \times n$ matrix and AB is an $m \times m$ matrix,

$$t^m \det(tI - BA) = t^n \det(tI - AB)$$

and so $\det(tI - BA) = p_{BA}(t) = t^{n-m}\det(tI - AB) = t^{n-m} p_{AB}(t)$. ■

3.6 Exercises

1. Let $m < n$ and let A be an $m \times n$ matrix. Show that A is **not** one to one. **Hint:** Consider the $n \times n$ matrix A_1 which is of the form

$$A_1 \equiv \begin{pmatrix} A \\ 0 \end{pmatrix}$$

where the 0 denotes an $(n - m) \times n$ matrix of zeros. Thus $\det A_1 = 0$ and so A_1 is not one to one. Now observe that $A_1 \mathbf{x}$ is the vector,

$$A_1 \mathbf{x} = \begin{pmatrix} A\mathbf{x} \\ \mathbf{0} \end{pmatrix}$$

which equals zero if and only if $A\mathbf{x} = \mathbf{0}$.

2. Let $\mathbf{v}_1, \cdots, \mathbf{v}_n$ be vectors in $\mathbb{F}^n$ and let $M(\mathbf{v}_1, \cdots, \mathbf{v}_n)$ denote the matrix whose i^{th} column equals $\mathbf{v}_i$. Define

$$d(\mathbf{v}_1, \cdots, \mathbf{v}_n) \equiv \det(M(\mathbf{v}_1, \cdots, \mathbf{v}_n)).$$

Prove that d is linear in each variable, (multilinear), that

$$d(\mathbf{v}_1, \cdots, \mathbf{v}_i, \cdots, \mathbf{v}_j, \cdots, \mathbf{v}_n) = -d(\mathbf{v}_1, \cdots, \mathbf{v}_j, \cdots, \mathbf{v}_i, \cdots, \mathbf{v}_n), \tag{3.22}$$

and

$$d(\mathbf{e}_1, \cdots, \mathbf{e}_n) = 1 \tag{3.23}$$

where here $\mathbf{e}_j$ is the vector in $\mathbb{F}^n$ which has a zero in every position except the j^{th}

position in which it has a one.

3. Suppose $f : \mathbb{F}^n \times \cdots \times \mathbb{F}^n \to \mathbb{F}$ satisfies 3.22 and 3.23 and is linear in each variable. Show that $f = d$.

4. Show that if you replace a row (column) of an $n \times n$ matrix A with itself added to some multiple of another row (column) then the new matrix has the same determinant as the original one.

5. Use the result of Problem 4 to evaluate by hand the determinant

$$\det \begin{pmatrix} 1 & 2 & 3 & 2 \\ -6 & 3 & 2 & 3 \\ 5 & 2 & 2 & 3 \\ 3 & 4 & 6 & 4 \end{pmatrix}.$$

6. Find the inverse if it exists of the matrix

$$\begin{pmatrix} e^t & \cos t & \sin t \\ e^t & -\sin t & \cos t \\ e^t & -\cos t & -\sin t \end{pmatrix}.$$

7. Let $Ly = y^{(n)} + a_{n-1}(x) y^{(n-1)} + \cdots + a_1(x) y' + a_0(x) y$ where the a_i are given continuous functions defined on an interval, (a,b) and y is some function which has n derivatives so it makes sense to write Ly. Suppose $Ly_k = 0$ for $k = 1, 2, \cdots, n$. The Wronskian of these functions, y_i is defined as

$$W(y_1, \cdots, y_n)(x) \equiv \det \begin{pmatrix} y_1(x) & \cdots & y_n(x) \\ y_1'(x) & \cdots & y_n'(x) \\ \vdots & & \vdots \\ y_1^{(n-1)}(x) & \cdots & y_n^{(n-1)}(x) \end{pmatrix}$$

Show that for $W(x) = W(y_1, \cdots, y_n)(x)$ to save space,

$$W'(x) = \det \begin{pmatrix} y_1(x) & \cdots & y_n(x) \\ \vdots & \cdots & \vdots \\ y_1^{(n-2)}(x) & & y_n^{(n-2)}(x) \\ y_1^{(n)}(x) & \cdots & y_n^{(n)}(x) \end{pmatrix}.$$

Now use the differential equation, $Ly = 0$ which is satisfied by each of these functions, y_i and properties of determinants presented above to verify that $W' + a_{n-1}(x) W = 0$. Give an explicit solution of this linear differential equation, Abel's formula, and use your answer to verify that the Wronskian of these solutions to the equation, $Ly = 0$ either vanishes identically on (a,b) or never.

8. Two $n \times n$ matrices, A and B, are similar if $B = S^{-1}AS$ for some invertible $n \times n$ matrix S. Show that if two matrices are similar, they have the same characteristic polynomials. The characteristic polynomial of A is $\det(\lambda I - A)$.

9. Suppose the characteristic polynomial of an $n \times n$ matrix A is of the form

$$t^n + a_{n-1}t^{n-1} + \cdots + a_1 t + a_0$$

and that $a_0 \neq 0$. Find a formula A^{-1} in terms of powers of the matrix A. Show that A^{-1} exists if and only if $a_0 \neq 0$. In fact, show that $a_0 = (-1)^n \det(A)$.

10. ↑Letting $p(t)$ denote the characteristic polynomial of A, show that $p_\varepsilon(t) \equiv p(t - \varepsilon)$ is the characteristic polynomial of $A + \varepsilon I$. Then show that if $\det(A) = 0$, it follows that $\det(A + \varepsilon I) \neq 0$ whenever $|\varepsilon|$ is sufficiently small.

11. In constitutive modeling of the stress and strain tensors, one sometimes considers sums of the form $\sum_{k=0}^{\infty} a_k A^k$ where A is a 3×3 matrix. Show using the Cayley Hamilton theorem that if such a thing makes any sense, you can always obtain it as a finite sum having no more than n terms.

12. Recall you can find the determinant from expanding along the j^{th} column.

$$\det(A) = \sum_i A_{ij} (\text{cof}(A))_{ij}$$

Think of $\det(A)$ as a function of the entries, A_{ij}. Explain why the ij^{th} cofactor is really just

$$\frac{\partial \det(A)}{\partial A_{ij}}.$$

13. Let U be an open set in $\mathbb{R}^n$ and let $\mathbf{g} : U \to \mathbb{R}^n$ be such that all the first partial derivatives of all components of $\mathbf{g}$ exist and are continuous. Under these conditions form the matrix $D\mathbf{g}(\mathbf{x})$ given by

$$D\mathbf{g}(\mathbf{x})_{ij} \equiv \frac{\partial g_i(\mathbf{x})}{\partial x_j} \equiv g_{i,j}(\mathbf{x})$$

The best kept secret in calculus courses is that the linear transformation determined by this matrix $D\mathbf{g}(\mathbf{x})$ is called the derivative of $\mathbf{g}$ and is the correct generalization of the concept of derivative of a function of one variable. Suppose the second partial derivatives also exist and are continuous. Then show that

$$\sum_j (\operatorname{cof}(D\mathbf{g}))_{ij,j} = 0.$$

Hint: First explain why $\sum_i g_{i,k} \operatorname{cof}(D\mathbf{g})_{ij} = \delta_{jk} \det(D\mathbf{g})$. Next differentiate with respect to x_j and sum on j using the equality of mixed partial derivatives. Assume $\det(D\mathbf{g}) \neq 0$ to prove the identity in this special case. Then explain using Problem 10 why there exists a sequence $\varepsilon_k \to 0$ such that for $\mathbf{g}_{\varepsilon_k}(\mathbf{x}) \equiv \mathbf{g}(\mathbf{x}) + \varepsilon_k \mathbf{x}$, $\det(D\mathbf{g}_{\varepsilon_k}) \neq 0$ and so the identity holds for $\mathbf{g}_{\varepsilon_k}$. Then take a limit to get the desired result in general. This is an extremely important identity which has surprising implications. One can build degree theory on it for example. It also leads to simple proofs of the Brouwer fixed point theorem from topology.

14. A determinant of the form

$$\begin{vmatrix} 1 & 1 & \cdots & 1 \\ a_0 & a_1 & \cdots & a_n \\ a_0^2 & a_1^2 & \cdots & a_n^2 \\ \vdots & \vdots & & \vdots \\ a_0^{n-1} & a_1^{n-1} & \cdots & a_n^{n-1} \\ a_0^n & a_1^n & \cdots & a_n^n \end{vmatrix}$$

is called a Vandermonde determinant. Show this determinant equals

$$\prod_{0 \leq i < j \leq n} (a_j - a_i)$$

By this is meant to take the product of all terms of the form $(a_j - a_i)$ such that $j > i$. **Hint:** Show it works if $n = 1$ so you are looking at

$$\begin{vmatrix} 1 & 1 \\ a_0 & a_1 \end{vmatrix}$$

Then suppose it holds for $n - 1$ and consider the case n. Consider the polynomial in $t, p(t)$ which is obtained from the above by replacing the last column with the column

$$\begin{pmatrix} 1 & t & \cdots & t^n \end{pmatrix}^T.$$

Explain why $p(a_j) = 0$ for $i = 0, \cdots, n-1$. Explain why

$$p(t) = c \prod_{i=0}^{n-1} (t - a_i).$$

Of course c is the coefficient of t^n. Find this coefficient from the above description of $p(t)$ and the induction hypothesis. Then plug in $t = a_n$ and observe you have the formula valid for n.

Row Operations

4.1 Elementary Matrices

The elementary matrices result from doing a row operation to the identity matrix.

Definition 4.1.1 *The row operations consist of the following*

1. *Switch two rows.*
2. *Multiply a row by a nonzero number.*
3. *Replace a row by a multiple of another row added to it.*

The elementary matrices are given in the following definition.

Definition 4.1.2 *The elementary matrices consist of those matrices which result by applying a row operation to an identity matrix. Those which involve switching rows of the identity are called permutation matrices. More generally, if* $(i_1, i_2, \cdots, i_n)$ *is a permutation, a matrix which has a 1 in the* i_k *position in row* k *and zero in every other position of that row is called a permutation matrix. Thus each permutation corresponds to a unique permutation matrix.*

As an example of why these elementary matrices are interesting, consider the following.

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a & b & c & d \\ x & y & z & w \\ f & g & h & i \end{pmatrix} = \begin{pmatrix} x & y & z & w \\ a & b & c & d \\ f & g & h & i \end{pmatrix}$$

A 3×4 matrix was multiplied on the left by an elementary matrix which was obtained from row operation 1 applied to the identity matrix. This resulted in applying the operation 1 to the given matrix. This is what happens in general.

Now consider what these elementary matrices look like. First consider the one which involves switching row i and row j where $i < j$. This matrix is of the form

$$\begin{pmatrix} 1 & & & & & & 0 \\ & \ddots & & & & & \\ & & 0 & \cdots & 1 & & \\ & & \vdots & & \vdots & & \\ & & 1 & \cdots & 0 & & \\ & & & & & \ddots & \\ 0 & & & & & & 1 \end{pmatrix}$$

The two exceptional rows are shown. The i^{th} row was the j^{th} and the j^{th} row was the i^{th} in the identity matrix. Now consider what this does to a column vector.

$$\begin{pmatrix} 1 & & & & & & 0 \\ & \ddots & & & & & \\ & & 0 & \cdots & 1 & & \\ & & \vdots & & \vdots & & \\ & & 1 & \cdots & 0 & & \\ & & & & & \ddots & \\ 0 & & & & & & 1 \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_i \\ \vdots \\ v_j \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} v_1 \\ \vdots \\ v_j \\ \vdots \\ v_i \\ \vdots \\ v_n \end{pmatrix}$$

Now denote by P^{ij} the elementary matrix which comes from the identity from switching rows i and j. From what was just explained consider multiplication on the left by this elementary matrix.

$$P^{ij} \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ \vdots & \vdots & & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{ip} \\ \vdots & \vdots & & \vdots \\ a_{j1} & a_{j2} & \cdots & a_{jp} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{np} \end{pmatrix}$$

From the way you multiply matrices this is a matrix which has the indicated columns.

$$\left(P^{ij} \begin{pmatrix} a_{11} \\ \vdots \\ a_{i1} \\ \vdots \\ a_{j1} \\ \vdots \\ a_{n1} \end{pmatrix}, P^{ij} \begin{pmatrix} a_{12} \\ \vdots \\ a_{i2} \\ \vdots \\ a_{j2} \\ \vdots \\ a_{n2} \end{pmatrix}, \cdots, P^{ij} \begin{pmatrix} a_{1p} \\ \vdots \\ a_{ip} \\ \vdots \\ a_{jp} \\ \vdots \\ a_{np} \end{pmatrix} \right)$$

$$= \left(\begin{pmatrix} a_{11} \\ \vdots \\ a_{j1} \\ \vdots \\ a_{i1} \\ \vdots \\ a_{n1} \end{pmatrix}, \begin{pmatrix} a_{12} \\ \vdots \\ a_{j2} \\ \vdots \\ a_{i2} \\ \vdots \\ a_{n2} \end{pmatrix}, \cdots, \begin{pmatrix} a_{1p} \\ \vdots \\ a_{jp} \\ \vdots \\ a_{ip} \\ \vdots \\ a_{np} \end{pmatrix} \right)$$

$$= \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ \vdots & \vdots & & \vdots \\ a_{j1} & a_{j2} & \cdots & a_{jp} \\ \vdots & \vdots & & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{ip} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{np} \end{pmatrix}$$

This has established the following lemma.

Lemma 4.1.3 *Let P^{ij} denote the elementary matrix which involves switching the i^{th} and the j^{th} rows. Then*

$$P^{ij}A = B$$

where B is obtained from A by switching the i^{th} and the j^{th} rows.

As a consequence of the above lemma, if you have any permutation $(i_1, \cdots, i_n)$, it follows from Lemma 3.3.2 that the corresponding permutation matrix can be obtained by multiplying finitely many permutation matrices, each of which switch only two rows. Now every such permutation matrix in which only two rows are switched has determinant -1. Therefore, the determinant of the permutation matrix for $(i_1, \cdots, i_n)$ equals $(-1)^p$ where the given permutation can be obtained by making p switches. Now p is not unique. There are many ways to make switches and end up with a given permutation, but what this shows is that the total number of switches is either always odd or always even. That is, you could not obtain a given permutation by making $2m$ switches and $2k+1$ switches. A permutation is said to be even if p is even and odd if p is odd. This is an interesting result in abstract algebra which is obtained very easily from a consideration of elementary matrices and of course the theory of the determinant. Also, this shows that the composition of permutations corresponds to the product of the corresponding permutation matrices.

To see permutations considered more directly in the context of group theory, you should see a good abstract algebra book such as [17] or [13].

Next consider the row operation which involves multiplying the i^{th} row by a nonzero constant, c. The elementary matrix which results from applying this operation to the i^{th} row of the identity matrix is of the form

$$\begin{pmatrix} 1 & & & & 0 \\ & \ddots & & & \\ & & c & & \\ & & & \ddots & \\ 0 & & & & 1 \end{pmatrix}$$

Now consider what this does to a column vector.

$$\begin{pmatrix} 1 & & & & 0 \\ & \ddots & & & \\ & & c & & \\ & & & \ddots & \\ 0 & & & & 1 \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_i \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} v_1 \\ \vdots \\ cv_i \\ \vdots \\ v_n \end{pmatrix}$$

Denote by $E(c,i)$ this elementary matrix which multiplies the i^{th} row of the identity by the nonzero constant, c. Then from what was just discussed and the way matrices are multiplied,

$$E(c,i)\begin{pmatrix} a_{11} & a_{12} & \cdots & \cdots & a_{1p} \\ \vdots & \vdots & & & \vdots \\ a_{i1} & a_{i2} & \cdots & \cdots & a_{ip} \\ \vdots & \vdots & & & \vdots \\ a_{n1} & a_{n2} & \cdots & \cdots & a_{np} \end{pmatrix}$$

equals a matrix having the columns indicated below.

$$= \left(E(c,i) \begin{pmatrix} a_{11} \\ \vdots \\ a_{i1} \\ \vdots \\ a_{n1} \end{pmatrix}, E(c,i) \begin{pmatrix} a_{12} \\ \vdots \\ a_{i2} \\ \vdots \\ a_{n2} \end{pmatrix}, \cdots, E(c,i) \begin{pmatrix} a_{1p} \\ \vdots \\ a_{ip} \\ \vdots \\ a_{np} \end{pmatrix} \right)$$

$$= \begin{pmatrix} a_{11} & a_{12} & \cdots & \cdots & a_{1p} \\ \vdots & \vdots & & & \vdots \\ ca_{i1} & ca_{i2} & \cdots & \cdots & ca_{ip} \\ \vdots & \vdots & & & \vdots \\ a_{n1} & a_{n2} & \cdots & \cdots & a_{np} \end{pmatrix}$$

This proves the following lemma.

Lemma 4.1.4 *Let $E(c,i)$ denote the elementary matrix corresponding to the row operation in which the i^{th} row is multiplied by the nonzero constant, c. Thus $E(c,i)$ involves multiplying the i^{th} row of the identity matrix by c. Then*

$$E(c,i)\,A = B$$

where B is obtained from A by multiplying the i^{th} row of A by c.

Finally consider the third of these row operations. Denote by $E(c \times i + j)$ the elementary matrix which replaces the j^{th} row with itself added to c times the i^{th} row added to it. In case $i < j$ this will be of the form

$$\begin{pmatrix} 1 & & & & & 0 \\ & \ddots & & & & \\ & & 1 & & & \\ & & \vdots & \ddots & & \\ & & c & \cdots & 1 & \\ & & & & & \ddots & \\ 0 & & & & & 1 \end{pmatrix}$$

Now consider what this does to a column vector.

$$\begin{pmatrix} 1 & & & & & 0 \\ & \ddots & & & & \\ & & 1 & & & \\ & & \vdots & \ddots & & \\ & & c & \cdots & 1 & \\ & & & & \ddots & \\ 0 & & & & & 1 \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_i \\ \vdots \\ v_j \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} v_1 \\ \vdots \\ v_i \\ \vdots \\ cv_i + v_j \\ \vdots \\ v_n \end{pmatrix}$$

Now from this and the way matrices are multiplied,

$$E(c \times i + j) \begin{pmatrix} a_{11} & a_{12} & \cdots & \cdots & \cdots & \cdots & a_{1p} \\ \vdots & \vdots & & & & & \vdots \\ a_{i1} & a_{i2} & \cdots & \cdots & \cdots & \cdots & a_{ip} \\ \vdots & \vdots & & & & & \vdots \\ a_{j2} & a_{j2} & \cdots & \cdots & \cdots & \cdots & a_{jp} \\ \vdots & \vdots & & & & & \vdots \\ a_{n1} & a_{n2} & \cdots & \cdots & \cdots & \cdots & a_{np} \end{pmatrix}$$

equals a matrix of the following form having the indicated columns.

$$\left(E(c \times i + j) \begin{pmatrix} a_{11} \\ \vdots \\ a_{i1} \\ \vdots \\ a_{j2} \\ \vdots \\ a_{n1} \end{pmatrix}, E(c \times i + j) \begin{pmatrix} a_{12} \\ \vdots \\ a_{i2} \\ \vdots \\ a_{j2} \\ \vdots \\ a_{n2} \end{pmatrix}, \cdots E(c \times i + j) \begin{pmatrix} a_{1p} \\ \vdots \\ a_{ip} \\ \vdots \\ a_{jp} \\ \vdots \\ a_{np} \end{pmatrix} \right)$$

$$= \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ \vdots & \vdots & & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{ip} \\ \vdots & \vdots & & \vdots \\ a_{j2} + ca_{i1} & a_{j2} + ca_{i2} & \cdots & a_{jp} + ca_{ip} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{np} \end{pmatrix}$$

The case where $i > j$ is handled similarly. This proves the following lemma.

Lemma 4.1.5 *Let $E(c \times i + j)$ denote the elementary matrix obtained from I by replacing the j^{th} row with c times the i^{th} row added to it. Then*

$$E(c \times i + j) A = B$$

where B is obtained from A by replacing the j^{th} row of A with itself added to c times the i^{th} row of A.

The next theorem is the main result.

Theorem 4.1.6 *To perform any of the three row operations on a matrix A it suffices to do the row operation on the identity matrix obtaining an elementary matrix E and then take the product, EA. Furthermore, each elementary matrix is invertible and its inverse is an elementary matrix.*

Proof: The first part of this theorem has been proved in Lemmas 4.1.3 - 4.1.5. It only remains to verify the claim about the inverses. Consider first the elementary matrices corresponding to row operation of type three.

$$E(-c \times i + j) E(c \times i + j) = I$$

This follows because the first matrix takes c times row i in the identity and adds it to row j. When multiplied on the left by $E(-c \times i + j)$ it follows from the first part of this theorem that you take the i^{th} row of $E(c \times i + j)$ which coincides with the i^{th} row of I since that row was not changed, multiply it by $-c$ and add to the j^{th} row of $E(c \times i + j)$ which was the j^{th} row of I added to c times the i^{th} row of I. Thus $E(-c \times i + j)$ multiplied on the left, undoes the row operation which resulted in $E(c \times i + j)$. The same argument applied to the product

$$E(c \times i + j)\, E(-c \times i + j)$$

replacing c with $-c$ in the argument yields that this product is also equal to I. Therefore, $E(c \times i + j)^{-1} = E(-c \times i + j)$.

Similar reasoning shows that for $E(c, i)$ the elementary matrix which comes from multiplying the i^{th} row by the nonzero constant, c,

$$E(c, i)^{-1} = E\left(c^{-1}, i\right).$$

Finally, consider P^{ij} which involves switching the i^{th} and the j^{th} rows.

$$P^{ij} P^{ij} = I$$

because by the first part of this theorem, multiplying on the left by P^{ij} switches the i^{th} and j^{th} rows of P^{ij} which was obtained from switching the i^{th} and j^{th} rows of the identity. First you switch them to get P^{ij} and then you multiply on the left by P^{ij} which switches these rows again and restores the identity matrix. Thus $\left(P^{ij}\right)^{-1} = P^{ij}$. ■

4.2 The Rank Of A Matrix

Recall the following definition of rank of a matrix.

Definition 4.2.1 *A submatrix of a matrix A is the rectangular array of numbers obtained by deleting some rows and columns of A. Let A be an $m \times n$ matrix. The* ***determinant rank*** *of the matrix equals r where r is the largest number such that some $r \times r$ submatrix of A has a non zero determinant. The* ***row rank*** *is defined to be the dimension of the span of the rows. The* ***column rank*** *is defined to be the dimension of the span of the columns. The rank of A is denoted as* $\operatorname{rank}(A)$.

The following theorem is proved in the section on the theory of the determinant and is restated here for convenience.

Theorem 4.2.2 *Let A be an $m \times n$ matrix. Then the row rank, column rank and determinant rank are all the same.*

So how do you find the rank? It turns out that row operations are the key to the practical computation of the rank of a matrix.

In rough terms, the following lemma states that **linear relationships** between columns in a matrix are preserved by row operations.

Lemma 4.2.3 *Let B and A be two $m \times n$ matrices and suppose B results from a row operation applied to A. Then the k^{th} column of B is a linear combination of the $i_1, \cdots, i_r$ columns of B if and only if the k^{th} column of A is a linear combination of the $i_1, \cdots, i_r$ columns of A. Furthermore, the scalars in the linear combination are the same. (The linear relationship between the k^{th} column of A and the $i_1, \cdots, i_r$ columns of A is the same as the linear relationship between the k^{th} column of B and the $i_1, \cdots, i_r$ columns of B.)*

Proof: Let A equal the following matrix in which the $\mathbf{a}_k$ are the columns

$$\begin{pmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{pmatrix}$$

and let B equal the following matrix in which the columns are given by the $\mathbf{b}_k$

$$\begin{pmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_n \end{pmatrix}$$

Then by Theorem 4.1.6 on Page 142 $\mathbf{b}_k = E\mathbf{a}_k$ where E is an elementary matrix. Suppose then that one of the columns of A is a linear combination of some other columns of A. Say

$$\mathbf{a}_k = \sum_{r \in S} c_r \mathbf{a}_r.$$

Then multiplying by E,

$$\mathbf{b}_k = E\mathbf{a}_k = \sum_{r \in S} c_r E\mathbf{a}_r = \sum_{r \in S} c_r \mathbf{b}_r. \blacksquare$$

Corollary 4.2.4 *Let A and B be two $m \times n$ matrices such that B is obtained by applying a row operation to A. Then the two matrices have the same rank.*

Proof: Lemma 4.2.3 says the linear relationships are the same between the columns of A and those of B. Therefore, the column rank of the two matrices is the same. $\blacksquare$

This suggests that to find the rank of a matrix, one should do row operations until a matrix is obtained in which its rank is obvious.

Example 4.2.5 *Find the rank of the following matrix and identify columns whose linear combinations yield all the other columns.*

$$\begin{pmatrix} 1 & 2 & 1 & 3 & 2 \\ 1 & 3 & 6 & 0 & 2 \\ 3 & 7 & 8 & 6 & 6 \end{pmatrix} \tag{4.1}$$

Take (-1) times the first row and add to the second and then take (-3) times the first row and add to the third. This yields

$$\begin{pmatrix} 1 & 2 & 1 & 3 & 2 \\ 0 & 1 & 5 & -3 & 0 \\ 0 & 1 & 5 & -3 & 0 \end{pmatrix}$$

By the above corollary, this matrix has the same rank as the first matrix. Now take (-1) times the second row and add to the third row yielding

$$\begin{pmatrix} 1 & 2 & 1 & 3 & 2 \\ 0 & 1 & 5 & -3 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

At this point it is clear the rank is 2. This is because every column is in the span of the first two and these first two columns are linearly independent.

Example 4.2.6 *Find the rank of the following matrix and identify columns whose linear combinations yield all the other columns.*

$$\begin{pmatrix} 1 & 2 & 1 & 3 & 2 \\ 1 & 2 & 6 & 0 & 2 \\ 3 & 6 & 8 & 6 & 6 \end{pmatrix} \tag{4.2}$$

Take (-1) times the first row and add to the second and then take (-3) times the first row and add to the last row. This yields

$$\begin{pmatrix} 1 & 2 & 1 & 3 & 2 \\ 0 & 0 & 5 & -3 & 0 \\ 0 & 0 & 5 & -3 & 0 \end{pmatrix}$$

Now multiply the second row by $1/5$ and add 5 times it to the last row.

$$\begin{pmatrix} 1 & 2 & 1 & 3 & 2 \\ 0 & 0 & 1 & -3/5 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Add (-1) times the second row to the first.

$$\begin{pmatrix} 1 & 2 & 0 & \frac{18}{5} & 2 \\ 0 & 0 & 1 & -3/5 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{4.3}$$

It is now clear the rank of this matrix is 2 because the first and third columns form a basis for the column space.

The matrix 4.3 is the row reduced echelon form for the matrix 4.2.

4.3 The Row Reduced Echelon Form

The following definition is for the row reduced echelon form of a matrix.

Definition 4.3.1 *Let $\mathbf{e}_i$ denote the column vector which has all zero entries except for the i^{th} slot which is one. An $m \times n$ matrix is said to be in row reduced echelon form if, in viewing successive columns from left to right, the first nonzero column encountered is $\mathbf{e}_1$ and if you have encountered $\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_k$, the next column is either $\mathbf{e}_{k+1}$ or is a linear combination of the vectors, $\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_k$.*

For example, here are some matrices which are in row reduced echelon form.

$$\begin{pmatrix} 0 & 1 & 3 & 0 & 3 \\ 0 & 0 & 0 & 1 & 5 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 3 & -11 & 0 \\ 0 & 1 & 4 & 4 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Theorem 4.3.2 *Let A be an $m \times n$ matrix. Then A has a row reduced echelon form determined by a simple process.*

Proof: Viewing the columns of A from left to right take the first nonzero column. Pick a nonzero entry in this column and switch the row containing this entry with the top row of A. Now divide this new top row by the value of this nonzero entry to get a 1 in this position and then use row operations to make all entries below this entry equal to zero. Thus the first nonzero column is now $\mathbf{e}_1$. Denote the resulting matrix by A_1. Consider the submatrix of A_1 to the right of this column and below the first row. Do exactly the same thing for it that was done for A. This time the $\mathbf{e}_1$ will refer to $\mathbb{F}^{m-1}$. Use this 1 and row operations to zero out every entry above it in the rows of A_1. Call the resulting matrix A_2. Thus A_2 satisfies the conditions of the above definition up to the column just encountered. Continue this way till every column has been dealt with and the result must be in row reduced echelon form. ■

The following diagram illustrates the above procedure. Say the matrix looked something like the following.

$$\begin{pmatrix} 0 & * & * & * & * & * & * \\ 0 & * & * & * & * & * & * \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & * & * & * & * & * & * \end{pmatrix}$$

First step would yield something like

$$\begin{pmatrix} 0 & 1 & * & * & * & * & * \\ 0 & 0 & * & * & * & * & * \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & * & * & * & * & * \end{pmatrix}$$

For the second step you look at the lower right corner as described,

$$\begin{pmatrix} * & * & * & * & * \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ * & * & * & * & * \end{pmatrix}$$

and if the first column consists of all zeros but the next one is not all zeros, you would get something like this.

$$\begin{pmatrix} 0 & 1 & * & * & * \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & * & * & * \end{pmatrix}$$

Thus, after zeroing out the term in the top row above the 1, you get the following for the next step in the computation of the row reduced echelon form for the original matrix.

$$\begin{pmatrix} 0 & 1 & * & 0 & * & * & * \\ 0 & 0 & 0 & 1 & * & * & * \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & * & * & * \end{pmatrix}.$$

Next you look at the lower right matrix below the top two rows and to the right of the first four columns and repeat the process.

Definition 4.3.3 *The first pivot column of A is the first nonzero column of A. The next pivot column is the first column after this which is not a linear combination of the columns to its left. The third pivot column is the next column after this which is not a linear combination of those columns to its left, and so forth. Thus by Lemma 4.2.3 if a pivot column occurs as the j^{th} column from the left, it follows that in the row reduced echelon form there will be one of the $\mathbf{e}_k$ as the j^{th} column.*

There are three choices for row operations at each step in the above theorem. A natural question is whether the same row reduced echelon matrix always results in the end from following the above algorithm applied in any way. The next corollary says this is the case.

Definition 4.3.4 *Two matrices are said to be **row equivalent** if one can be obtained from the other by a sequence of row operations.*

Since every row operation can be obtained by multiplication on the left by an elementary matrix and since each of these elementary matrices has an inverse which is also an elementary matrix, it follows that row equivalence is a similarity relation. Thus one can classify matrices according to which similarity class they are in. Later in the book, another more profound way of classifying matrices will be presented.

It has been shown above that every matrix is row equivalent to one which is in row reduced echelon form. Note

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n$$

so to say two column vectors are equal is to say they are the same linear combination of the special vectors $\mathbf{e}_j$.

Corollary 4.3.5 *The row reduced echelon form is unique. That is if B, C are two matrices in row reduced echelon form and both are row equivalent to A, then $B = C$.*

Proof: Suppose B and C are both row reduced echelon forms for the matrix A. Then they clearly have the same zero columns since row operations leave zero columns unchanged. If B has the sequence $\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_r$ occurring for the first time in the positions, $i_1, i_2, \cdots, i_r$, the description of the row reduced echelon form means that each of these columns is **not** a linear combination of the preceding columns. Therefore, by Lemma 4.2.3, the same is true of the columns in positions $i_1, i_2, \cdots, i_r$ for C. It follows from the description of the row reduced echelon form, that $\mathbf{e}_1, \cdots, \mathbf{e}_r$ occur respectively for the first time in columns $i_1, i_2, \cdots, i_r$ for C. Thus B, C have the same columns in these positions. By Lemma 4.2.3, the other columns in the two matrices are linear combinations, involving the **same scalars**, of the columns in the $i_1, \cdots, i_k$ position. Thus each column of B is identical to the corresponding column in C. ■

The above corollary shows that you can determine whether two matrices are row equivalent by simply checking their row reduced echelon forms. The matrices are row equivalent if and only if they have the same row reduced echelon form.

The following corollary follows.

Corollary 4.3.6 *Let A be an $m \times n$ matrix and let R denote the row reduced echelon form obtained from A by row operations. Then there exists a sequence of elementary matrices, $E_1, \cdots, E_p$ such that*

$$(E_p E_{p-1} \cdots E_1) A = R.$$

Proof: This follows from the fact that row operations are equivalent to multiplication on the left by an elementary matrix. ■

Corollary 4.3.7 *Let A be an invertible $n \times n$ matrix. Then A equals a finite product of elementary matrices.*

Proof: Since A^{-1} is given to exist, it follows A must have rank n because by Theorem 3.3.18 $\det(A) \neq 0$ which says the determinant rank and hence the column rank of A is n and so the row reduced echelon form of A is I because the columns of A form a linearly independent set. Therefore, by Corollary 4.3.6 there is a sequence of elementary matrices, $E_1, \cdots, E_p$ such that

$$(E_p E_{p-1} \cdots E_1) A = I.$$

But now multiply on the left on both sides by E_p^{-1} then by E_{p-1}^{-1} and then by E_{p-2}^{-1} etc. until you get

$$A = E_1^{-1} E_2^{-1} \cdots E_{p-1}^{-1} E_p^{-1}$$

and by Theorem 4.1.6 each of these in this product is an elementary matrix.

Corollary 4.3.8 *The rank of a matrix equals the number of nonzero pivot columns. Furthermore, every column is contained in the span of the pivot columns.*

Proof: Write the row reduced echelon form for the matrix. From Corollary 4.2.4 this row reduced matrix has the same rank as the original matrix. Deleting all the zero rows and all the columns in the row reduced echelon form which do not correspond to a pivot column, yields an $r \times r$ identity submatrix in which r is the number of pivot columns. Thus the rank is at least r.

From Lemma 4.2.3 every column of A is a linear combination of the pivot columns since this is true by definition for the row reduced echelon form. Therefore, the rank is no more than r. ■

Here is a fundamental observation related to the above.

Corollary 4.3.9 *Suppose A is an $m \times n$ matrix and that $m < n$. That is, the number of rows is less than the number of columns. Then one of the columns of A is a linear combination of the preceding columns of A.*

Proof: Since $m < n$, not all the columns of A can be pivot columns. That is, in the row reduced echelon form say $\mathbf{e}_i$ occurs for the first time at r_i where $r_1 < r_2 < \cdots < r_p$ where $p \leq m$. It follows since $m < n$, there exists some column in the row reduced echelon form which is a linear combination of the preceding columns. By Lemma 4.2.3 the same is true of the columns of A. ■

Definition 4.3.10 *Let A be an $m \times n$ matrix having rank, r. Then the nullity of A is defined to be $n - r$. Also define* $\ker(A) \equiv \{\mathbf{x} \in \mathbb{F}^n : A\mathbf{x} = \mathbf{0}\}$. *This is also denoted as $N(A)$.*

Observation 4.3.11 *Note that* $\ker(A)$ *is a subspace because if a, b are scalars and* $\mathbf{x}, \mathbf{y}$ *are vectors in* $\ker(A)$, *then*

$$A(a\mathbf{x} + b\mathbf{y}) = aA\mathbf{x} + bA\mathbf{y} = \mathbf{0} + \mathbf{0} = \mathbf{0}$$

Recall that the dimension of the column space of a matrix equals its rank and since the column space is just $A(\mathbb{F}^n)$, the rank is just the dimension of $A(\mathbb{F}^n)$. The next theorem shows that the nullity equals the dimension of $\ker(A)$.

Theorem 4.3.12 *Let A be an $m \times n$ matrix. Then* $\operatorname{rank}(A) + \dim(\ker(A)) = n$..

Proof: Since $\ker(A)$ is a subspace, there exists a basis for $\ker(A)$, $\{\mathbf{x}_1, \cdots, \mathbf{x}_k\}$. Also let $\{A\mathbf{y}_1, \cdots, A\mathbf{y}_l\}$ be a basis for $A(\mathbb{F}^n)$. Let $\mathbf{u} \in \mathbb{F}^n$. Then there exist unique scalars c_i such that

$$A\mathbf{u} = \sum_{i=1}^{l} c_i A\mathbf{y}_i$$

It follows that

$$A\left(\mathbf{u} - \sum_{i=1}^{l} c_i \mathbf{y}_i\right) = \mathbf{0}$$

and so the vector in parenthesis is in $\ker(A)$. Thus there exist unique b_j such that

$$\mathbf{u} = \sum_{i=1}^{l} c_i \mathbf{y}_i + \sum_{j=1}^{k} b_j \mathbf{x}_j$$

Since $\mathbf{u}$ was arbitrary, this shows $\{\mathbf{x}_1, \cdots, \mathbf{x}_k, \mathbf{y}_1, \cdots, \mathbf{y}_l\}$ spans $\mathbb{F}^n$. If these vectors are independent, then they will form a basis and the claimed equation will be obtained. Suppose then that

$$\sum_{i=1}^{l} c_i \mathbf{y}_i + \sum_{j=1}^{k} b_j \mathbf{x}_j = \mathbf{0}$$

Apply A to both sides. This yields

$$\sum_{i=1}^{l} c_i A \mathbf{y}_i = \mathbf{0}$$

and so each $c_i = 0$. Then the independence of the $\mathbf{x}_j$ imply each $b_j = 0$. ■

4.4 Rank And Existence Of Solutions To Linear Systems

Consider the linear system of equations,

$$A\mathbf{x} = \mathbf{b} \tag{4.4}$$

where A is an $m \times n$ matrix, $\mathbf{x}$ is a $n \times 1$ column vector, and $\mathbf{b}$ is an $m \times 1$ column vector. Suppose

$$A = \begin{pmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{pmatrix}$$

where the $\mathbf{a}_k$ denote the columns of A. Then $\mathbf{x} = (x_1, \cdots, x_n)^T$ is a solution of the system 4.4, if and only if

$$x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n = \mathbf{b}$$

which says that $\mathbf{b}$ is a vector in $\text{span}(\mathbf{a}_1, \cdots, \mathbf{a}_n)$. This shows that there exists a solution to the system, 4.4 if and only if $\mathbf{b}$ is contained in $\text{span}(\mathbf{a}_1, \cdots, \mathbf{a}_n)$. In words, there is a solution to 4.4 if and only if $\mathbf{b}$ is in the column space of A. In terms of rank, the following proposition describes the situation.

Proposition 4.4.1 *Let A be an $m \times n$ matrix and let $\mathbf{b}$ be an $m \times 1$ column vector. Then there exists a solution to 4.4 if and only if*

$$\text{rank}\begin{pmatrix} A & | & \mathbf{b} \end{pmatrix} = \text{rank}(A). \tag{4.5}$$

Proof: Place $\begin{pmatrix} A & | & \mathbf{b} \end{pmatrix}$ and A in row reduced echelon form, respectively B and C. If the above condition on rank is true, then both B and C have the same number of nonzero rows. In particular, you cannot have a row of the form

$$\begin{pmatrix} 0 & \cdots & 0 & \bigstar \end{pmatrix}$$

where $\bigstar \neq 0$ in B. Therefore, there will exist a solution to the system 4.4.

Conversely, suppose there exists a solution. This means there cannot be such a row in B described above. Therefore, B and C must have the same number of zero rows and so they have the same number of nonzero rows. Therefore, the rank of the two matrices in 4.5 is the same. ■

4.5 Fredholm Alternative

There is a very useful version of Proposition 4.4.1 known as the **Fredholm alternative**. I will only present this for the case of real matrices here. Later a much more elegant and general approach is presented which allows for the general case of complex matrices.

The following definition is used to state the Fredholm alternative.

Definition 4.5.1 *Let $S \subseteq \mathbb{R}^m$. Then $S^{\perp} \equiv \{\mathbf{z} \in \mathbb{R}^m : \mathbf{z} \cdot \mathbf{s} = 0$ for every $\mathbf{s} \in S\}$. The funny exponent, $\perp$ is called "perp".*

Now note

$$\ker\left(A^T\right) \equiv \left\{\mathbf{z} : A^T\mathbf{z} = \mathbf{0}\right\} = \left\{\mathbf{z} : \sum_{k=1}^{m} z_k \mathbf{a}_k = 0\right\}$$

Lemma 4.5.2 *Let A be a real $m \times n$ matrix, let $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^m$. Then*

$$(A\mathbf{x} \cdot \mathbf{y}) = \left(\mathbf{x} \cdot A^T\mathbf{y}\right)$$

Proof: This follows right away from the definition of the inner product and matrix multiplication.

$$(A\mathbf{x} \cdot \mathbf{y}) = \sum_{k,l} A_{kl} x_l y_k = \sum_{k,l} \left(A^T\right)_{lk} x_l y_k = \left(\mathbf{x} \cdot A^T\mathbf{y}\right). \blacksquare$$

Now it is time to state the Fredholm alternative. The first version of this is the following theorem.

Theorem 4.5.3 *Let A be a real $m \times n$ matrix and let $\mathbf{b} \in \mathbb{R}^m$. There exists a solution, $\mathbf{x}$ to the equation $A\mathbf{x} = \mathbf{b}$ if and only if $\mathbf{b} \in \ker\left(A^T\right)^{\perp}$.*

Proof: First suppose $\mathbf{b} \in \ker\left(A^T\right)^{\perp}$. Then this says that if $A^T\mathbf{x} = \mathbf{0}$, it follows that $\mathbf{b} \cdot \mathbf{x} = \mathbf{0}$. In other words, taking the transpose, if

$$\mathbf{x}^T A = \mathbf{0}, \text{ then } \mathbf{x}^T\mathbf{b} = 0.$$

Thus, if P is a product of elementary matrices such that PA is in row reduced echelon form, then if PA has a row of zeros, in the k^{th} position, then there is also a zero in the k^{th} position of $P\mathbf{b}$. Thus rank $\begin{pmatrix} A & | & \mathbf{b} \end{pmatrix} = \operatorname{rank}(A)$.By Proposition 4.4.1, there exists a solution, $\mathbf{x}$ to the system $A\mathbf{x} = \mathbf{b}$. It remains to go the other direction.

Let $\mathbf{z} \in \ker\left(A^T\right)$ and suppose $A\mathbf{x} = \mathbf{b}$. I need to verify $\mathbf{b} \cdot \mathbf{z} = 0$. By Lemma 4.5.2,

$$\mathbf{b} \cdot \mathbf{z} = A\mathbf{x} \cdot \mathbf{z} = \mathbf{x} \cdot A^T\mathbf{z} = \mathbf{x} \cdot \mathbf{0} = 0 \blacksquare$$

This implies the following corollary which is also called the Fredholm alternative. The "alternative" becomes more clear in this corollary.

Corollary 4.5.4 *Let A be an $m \times n$ matrix. Then A maps $\mathbb{R}^n$ onto $\mathbb{R}^m$ if and only if the only solution to $A^T\mathbf{x} = \mathbf{0}$ is $\mathbf{x} = \mathbf{0}$.*

Proof: If the only solution to $A^T\mathbf{x} = \mathbf{0}$ is $\mathbf{x} = \mathbf{0}$, then $\ker\left(A^T\right) = \{\mathbf{0}\}$ and so $\ker\left(A^T\right)^{\perp} = \mathbb{R}^m$ because every $\mathbf{b} \in \mathbb{R}^m$ has the property that $\mathbf{b} \cdot \mathbf{0} = 0$. Therefore, $A\mathbf{x} = \mathbf{b}$ has a solution for any $\mathbf{b} \in \mathbb{R}^m$ because the $\mathbf{b}$ for which there is a solution are those in $\ker\left(A^T\right)^{\perp}$ by Theorem 4.5.3. In other words, A maps $\mathbb{R}^n$ onto $\mathbb{R}^m$.

Conversely if A is onto, then by Theorem 4.5.3 every $\mathbf{b} \in \mathbb{R}^m$ is in $\ker\left(A^T\right)^{\perp}$ and so if $A^T\mathbf{x} = \mathbf{0}$, then $\mathbf{b} \cdot \mathbf{x} = 0$ for every $\mathbf{b}$. In particular, this holds for $\mathbf{b} = \mathbf{x}$. Hence if $A^T\mathbf{x} = \mathbf{0}$, then $\mathbf{x} = \mathbf{0}$. ■

Here is an amusing example.

Example 4.5.5 *Let A be an $m \times n$ matrix in which $m > n$. Then A cannot map onto $\mathbb{R}^m$.*

The reason for this is that A^T is an $n \times m$ where $m > n$ and so in the augmented matrix

$$\left(A^T|\mathbf{0}\right)$$

there must be some free variables. Thus there exists a nonzero vector $\mathbf{x}$ such that $A^T\mathbf{x} = \mathbf{0}$.

4.6 Exercises

1. Let $\{\mathbf{u}_1, \cdots, \mathbf{u}_n\}$ be vectors in $\mathbb{R}^n$. The parallelepiped determined by these vectors $P(\mathbf{u}_1, \cdots, \mathbf{u}_n)$ is defined as

$$P(\mathbf{u}_1, \cdots, \mathbf{u}_n) \equiv \left\{ \sum_{k=1}^{n} t_k \mathbf{u}_k : t_k \in [0, 1] \text{ for all } k \right\}.$$

   Now let A be an $n \times n$ matrix. Show that

$$\{A\mathbf{x} : \mathbf{x} \in P(\mathbf{u}_1, \cdots, \mathbf{u}_n)\}$$

   is also a parallelepiped.

2. In the context of Problem 1, draw $P(\mathbf{e}_1, \mathbf{e}_2)$ where $\mathbf{e}_1, \mathbf{e}_2$ are the standard basis vectors for $\mathbb{R}^2$. Thus $\mathbf{e}_1 = (1, 0), \mathbf{e}_2 = (0, 1)$. Now suppose

$$E = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

   where E is the elementary matrix which takes the third row and adds to the first. Draw

$$\{E\mathbf{x} : \mathbf{x} \in P(\mathbf{e}_1, \mathbf{e}_2)\}.$$

   In other words, draw the result of doing E to the vectors in $P(\mathbf{e}_1, \mathbf{e}_2)$. Next draw the results of doing the other elementary matrices to $P(\mathbf{e}_1, \mathbf{e}_2)$.

3. In the context of Problem 1, either draw or describe the result of doing elementary matrices to $P(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$. Describe geometrically the conclusion of Corollary 4.3.7.

4. Consider a permutation of $\{1, 2, \cdots, n\}$. This is an ordered list of numbers taken from this list with no repeats, $\{i_1, i_2, \cdots, i_n\}$. Define the permutation matrix $P(i_1, i_2, \cdots, i_n)$ as the matrix which is obtained from the identity matrix by placing the j^{th} column of I as the i_j^{th} column of $P(i_1, i_2, \cdots, i_n)$. Thus the 1 in the i_j^{th} column of this permutation matrix occurs in the j^{th} slot. What does this permutation matrix do to the column vector $(1, 2, \cdots, n)^T$?

5. $\uparrow$Consider the 3×3 permutation matrices. List all of them and then determine the dimension of their span. Recall that you can consider an $m \times n$ matrix as something in $\mathbb{F}^{nm}$.

6. Determine which matrices are in row reduced echelon form.

   (a) $\begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 7 \end{pmatrix}$

   (b) $\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{pmatrix}$

   (c) $\begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 5 \\ 0 & 0 & 1 & 2 & 0 & 4 \\ 0 & 0 & 0 & 0 & 1 & 3 \end{pmatrix}$

7. Row reduce the following matrices to obtain the row reduced echelon form. List the pivot columns in the original matrix.

   (a) $\begin{pmatrix} 1 & 2 & 0 & 3 \\ 2 & 1 & 2 & 2 \\ 1 & 1 & 0 & 3 \end{pmatrix}$

   (b) $\begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & -2 \\ 3 & 0 & 0 \\ 3 & 2 & 1 \end{pmatrix}$

   (c) $\begin{pmatrix} 1 & 2 & 1 & 3 \\ -3 & 2 & 1 & 0 \\ 3 & 2 & 1 & 1 \end{pmatrix}$

8. Find the rank and nullity of the following matrices. If the rank is r, identify r columns **in the original matrix** which have the property that every other column may be written as a linear combination of these.

   (a) $\begin{pmatrix} 0 & 1 & 0 & 2 & 1 & 2 & 2 \\ 0 & 3 & 2 & 12 & 1 & 6 & 8 \\ 0 & 1 & 1 & 5 & 0 & 2 & 3 \\ 0 & 2 & 1 & 7 & 0 & 3 & 4 \end{pmatrix}$

   (b) $\begin{pmatrix} 0 & 1 & 0 & 2 & 0 & 1 & 0 \\ 0 & 3 & 2 & 6 & 0 & 5 & 4 \\ 0 & 1 & 1 & 2 & 0 & 2 & 2 \\ 0 & 2 & 1 & 4 & 0 & 3 & 2 \end{pmatrix}$

   (c) $\begin{pmatrix} 0 & 1 & 0 & 2 & 1 & 1 & 2 \\ 0 & 3 & 2 & 6 & 1 & 5 & 1 \\ 0 & 1 & 1 & 2 & 0 & 2 & 1 \\ 0 & 2 & 1 & 4 & 0 & 3 & 1 \end{pmatrix}$

9. Find the rank of the following matrices. If the rank is r, identify r columns **in the original matrix** which have the property that every other column may be written as a linear combination of these. Also find a basis for the row and column spaces of the matrices.

(a) $\begin{pmatrix} 1 & 2 & 0 \\ 3 & 2 & 1 \\ 2 & 1 & 0 \\ 0 & 2 & 1 \end{pmatrix}$

(b) $\begin{pmatrix} 1 & 0 & 0 \\ 4 & 1 & 1 \\ 2 & 1 & 0 \\ 0 & 2 & 0 \end{pmatrix}$

(c) $\begin{pmatrix} 0 & 1 & 0 & 2 & 1 & 2 & 2 \\ 0 & 3 & 2 & 12 & 1 & 6 & 8 \\ 0 & 1 & 1 & 5 & 0 & 2 & 3 \\ 0 & 2 & 1 & 7 & 0 & 3 & 4 \end{pmatrix}$

(d) $\begin{pmatrix} 0 & 1 & 0 & 2 & 0 & 1 & 0 \\ 0 & 3 & 2 & 6 & 0 & 5 & 4 \\ 0 & 1 & 1 & 2 & 0 & 2 & 2 \\ 0 & 2 & 1 & 4 & 0 & 3 & 2 \end{pmatrix}$

(e) $\begin{pmatrix} 0 & 1 & 0 & 2 & 1 & 1 & 2 \\ 0 & 3 & 2 & 6 & 1 & 5 & 1 \\ 0 & 1 & 1 & 2 & 0 & 2 & 1 \\ 0 & 2 & 1 & 4 & 0 & 3 & 1 \end{pmatrix}$

10. Suppose A is an $m \times n$ matrix. Explain why the rank of A is always no larger than $\min(m, n)$.

11. Suppose A is an $m \times n$ matrix in which $m \leq n$. Suppose also that the rank of A equals m. Show that A maps $\mathbb{F}^n$ onto $\mathbb{F}^m$. **Hint:** The vectors $\mathbf{e}_1, \cdots, \mathbf{e}_m$ occur as columns in the row reduced echelon form for A.

12. Suppose A is an $m \times n$ matrix and that $m > n$. Show there exists $\mathbf{b} \in \mathbb{F}^m$ such that there is no solution to the equation

$$A\mathbf{x} = \mathbf{b}.$$

13. Suppose A is an $m \times n$ matrix in which $m \geq n$. Suppose also that the rank of A equals n. Show that A is one to one. **Hint:** If not, there exists a vector, $\mathbf{x} \neq \mathbf{0}$ such that $A\mathbf{x} = \mathbf{0}$, and this implies at least one column of A is a linear combination of the others. Show this would require the column rank to be less than n.

14. Explain why an $n \times n$ matrix A is both one to one and onto if and only if its rank is n.

15. Suppose A is an $m \times n$ matrix and $\{\mathbf{w}_1, \cdots, \mathbf{w}_k\}$ is a linearly independent set of vectors in $A(\mathbb{F}^n) \subseteq \mathbb{F}^m$. Suppose also that $A\mathbf{z}_i = \mathbf{w}_i$. Show that $\{\mathbf{z}_1, \cdots, \mathbf{z}_k\}$ is also linearly independent.

16. Show $\text{rank}(A + B) \leq \text{rank}(A) + \text{rank}(B)$.

17. Suppose A is an $m \times n$ matrix, $m \geq n$ and the columns of A are independent. Suppose also that $\{\mathbf{z}_1, \cdots, \mathbf{z}_k\}$ is a linearly independent set of vectors in $\mathbb{F}^n$. Show that $\{A\mathbf{z}_1, \cdots, A\mathbf{z}_k\}$ is linearly independent.

18. Suppose A is an $m \times n$ matrix and B is an $n \times p$ matrix. Show that

$$\dim(\ker(AB)) \leq \dim(\ker(A)) + \dim(\ker(B)).$$

Hint: Consider the subspace, $B(\mathbb{F}^p) \cap \ker(A)$ and suppose a basis for this subspace is $\{\mathbf{w}_1, \cdots, \mathbf{w}_k\}$. Now suppose $\{\mathbf{u}_1, \cdots, \mathbf{u}_r\}$ is a basis for $\ker(B)$. Let $\{\mathbf{z}_1, \cdots, \mathbf{z}_k\}$ be such that $B\mathbf{z}_i = \mathbf{w}_i$ and argue that

$$\ker(AB) \subseteq \operatorname{span}(\mathbf{u}_1, \cdots, \mathbf{u}_r, \mathbf{z}_1, \cdots, \mathbf{z}_k).$$

19. Let $m < n$ and let A be an $m \times n$ matrix. Show that A is **not** one to one.

20. Let A be an $m \times n$ real matrix and let $\mathbf{b} \in \mathbb{R}^m$. Show there exists a solution, $\mathbf{x}$ to the system

$$A^T A\mathbf{x} = A^T \mathbf{b}$$

Next show that if $\mathbf{x}, \mathbf{x}_1$ are two solutions, then $A\mathbf{x} = A\mathbf{x}_1$. **Hint:** First show that $\left(A^T A\right)^T = A^T A$. Next show if $\mathbf{x} \in \ker\left(A^T A\right)$, then $A\mathbf{x} = \mathbf{0}$. Finally apply the Fredholm alternative. Show $A^T\mathbf{b} \in \ker(A^T A)^{\perp}$. This will give existence of a solution.

21. Show that in the context of Problem 20 that if $\mathbf{x}$ is the solution there, then $|\mathbf{b} - A\mathbf{x}| \leq |\mathbf{b} - A\mathbf{y}|$ for every $\mathbf{y}$. Thus $A\mathbf{x}$ is the point of $A(\mathbb{R}^n)$ which is closest to $\mathbf{b}$ of every point in $A(\mathbb{R}^n)$. This is a solution to the least squares problem.

22. $\uparrow$Here is a point in $\mathbb{R}^4 : (1,2,3,4)^T$. Find the point in span $\left(\begin{pmatrix} 1 \\ 0 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 3 \\ 2 \end{pmatrix}\right)$ which is closest to the given point.

23. $\uparrow$Here is a point in $\mathbb{R}^4 : (1,2,3,4)^T$. Find the point on the plane described by $x + 2y - 4z + 4w = 0$ which is closest to the given point.

24. Suppose A, B are two invertible $n \times n$ matrices. Show there exists a sequence of row operations which when done to A yield B. **Hint:** Recall that every invertible matrix is a product of elementary matrices.

25. If A is invertible and $n \times n$ and B is $n \times p$, show that AB has the same null space as B and also the same rank as B.

26. Here are two matrices in row reduced echelon form

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \; B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

Does there exist a sequence of row operations which when done to A will yield B? Explain.

27. Is it true that an upper triagular matrix has rank equal to the number of nonzero entries down the main diagonal?

28. Let $\{\mathbf{v}_1, \cdots, \mathbf{v}_{n-1}\}$ be vectors in $\mathbb{F}^n$. Describe a systematic way to obtain a vector $\mathbf{v}_n$ which is perpendicular to each of these vectors. **Hint:** You might consider something like this

$$\det \begin{pmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_n \\ v_{11} & v_{12} & \cdots & v_{1n} \\ \vdots & \vdots & & \vdots \\ v_{(n-1)1} & v_{(n-1)2} & \cdots & v_{(n-1)n} \end{pmatrix}$$

where v_{ij} is the j^{th} entry of the vector $\mathbf{v}_i$. This is a lot like the cross product.

29. Let A be an $m \times n$ matrix. Then $\ker(A)$ is a subspace of $\mathbb{F}^n$. Is it true that every subspace of $\mathbb{F}^n$ is the kernel or null space of some matrix? Prove or disprove.

30. Let A be an $n \times n$ matrix and let P^{ij} be the permutation matrix which switches the i^{th} and j^{th} rows of the identity. Show that $P^{ij} A P^{ij}$ produces a matrix which is similar to A which switches the i^{th} and j^{th} entries on the main diagonal.

31. Recall the procedure for finding the inverse of a matrix on Page 62. It was shown that the procedure, when it works, finds the inverse of the matrix. Show that whenever the matrix has an inverse, the procedure works.

Some Factorizations

5.1 LU Factorization

An LU factorization of a matrix involves writing the given matrix as the product of a lower triangular matrix which has the main diagonal consisting entirely of ones, L, and an upper triangular matrix U in the indicated order. The L goes with "lower" and the U with "upper". It turns out many matrices can be written in this way and when this is possible, people get excited about slick ways of solving the system of equations, $A\mathbf{x} = \mathbf{y}$. The method lacks generality but is of interest just the same.

Example 5.1.1 *Can you write* $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ *in the form LU as just described?*

To do so you would need

$$\begin{pmatrix} 1 & 0 \\ x & 1 \end{pmatrix} \begin{pmatrix} a & b \\ 0 & c \end{pmatrix} = \begin{pmatrix} a & b \\ xa & xb + c \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Therefore, $b = 1$ and $a = 0$. Also, from the bottom rows, $xa = 1$ which can't happen and have $a = 0$. Therefore, you can't write this matrix in the form LU. It has no LU factorization. This is what I mean above by saying the method lacks generality.

Which matrices have an LU factorization? It turns out it is those whose row reduced echelon form can be achieved without switching rows and which only involve row operations of type 3 in which row j is replaced with a multiple of row i added to row j for $i < j$.

5.2 Finding An LU Factorization

There is a convenient procedure for finding an LU factorization. It turns out that it is only necessary to keep track of the **multipliers** which are used to row reduce to upper triangular form. This procedure is described in the following examples and is called the multiplier method. It is due to Dolittle.

Example 5.2.1 *Find an LU factorization for* $A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & -4 \\ 1 & 5 & 2 \end{pmatrix}$

Write the matrix next to the identity matrix as shown.

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & -4 \\ 1 & 5 & 2 \end{pmatrix}.$$

The process involves doing row operations to the matrix on the right while simultaneously updating successive columns of the matrix on the left. First take -2 times the first row and add to the second in the matrix on the right.

$$\begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 0 & -3 & -10 \\ 1 & 5 & 2 \end{pmatrix}$$

Note the method for updating the matrix on the left. The 2 in the second entry of the first column is there because -2 times the first row of A added to the second row of A produced a 0. Now replace the third row in the matrix on the right by -1 times the first row added to the third. Thus the next step is

$$\begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 0 & -3 & -10 \\ 0 & 3 & -1 \end{pmatrix}$$

Finally, add the second row to the bottom row and make the following changes

$$\begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 0 & -3 & -10 \\ 0 & 0 & -11 \end{pmatrix}.$$

At this point, stop because the matrix on the right is upper triangular. An LU factorization is the above.

The justification for this gimmick will be given later.

Example 5.2.2 *Find an LU factorization for* $A = \begin{pmatrix} 1 & 2 & 1 & 2 & 1 \\ 2 & 0 & 2 & 1 & 1 \\ 2 & 3 & 1 & 3 & 2 \\ 1 & 0 & 1 & 1 & 2 \end{pmatrix}$.

This time everything is done at once for a whole column. This saves trouble. First multiply the first row by (-1) and then add to the last row. Next take (-2) times the first and add to the second and then (-2) times the first and add to the third.

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 2 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 1 & 2 & 1 \\ 0 & -4 & 0 & -3 & -1 \\ 0 & -1 & -1 & -1 & 0 \\ 0 & -2 & 0 & -1 & 1 \end{pmatrix}.$$

This finishes the first column of L and the first column of U. Now take $-(1/4)$ times the second row in the matrix on the right and add to the third followed by $-(1/2)$ times the second added to the last.

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 2 & 1/4 & 1 & 0 \\ 1 & 1/2 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 1 & 2 & 1 \\ 0 & -4 & 0 & -3 & -1 \\ 0 & 0 & -1 & -1/4 & 1/4 \\ 0 & 0 & 0 & 1/2 & 3/2 \end{pmatrix}$$

This finishes the second column of L as well as the second column of U. Since the matrix on the right is upper triangular, stop. The LU factorization has now been obtained. This technique is called Dolittle's method. ►►

This process is entirely typical of the general case. The matrix U is just the first upper triangular matrix you come to in your quest for the row reduced echelon form using only the row operation which involves replacing a row by itself added to a multiple of another row. The matrix L is what you get by updating the identity matrix as illustrated above.

You should note that for a square matrix, the number of row operations necessary to reduce to LU form is about half the number needed to place the matrix in row reduced echelon form. This is why an LU factorization is of interest in solving systems of equations.

5.3 Solving Linear Systems Using An LU Factorization

The reason people care about the LU factorization is it allows the quick solution of systems of equations. Here is an example.

Example 5.3.1 *Suppose you want to find the solutions to* $\begin{pmatrix} 1 & 2 & 3 & 2 \\ 4 & 3 & 1 & 1 \\ 1 & 2 & 3 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$.

Of course one way is to write the augmented matrix and grind away. However, this involves more row operations than the computation of an LU factorization and it turns out that an LU factorization can give the solution quickly. Here is how. The following is an LU factorization for the matrix.

$$\begin{pmatrix} 1 & 2 & 3 & 2 \\ 4 & 3 & 1 & 1 \\ 1 & 2 & 3 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 4 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 2 \\ 0 & -5 & -11 & -7 \\ 0 & 0 & 0 & -2 \end{pmatrix}.$$

Let $U\mathbf{x} = \mathbf{y}$ and consider $L\mathbf{y} = \mathbf{b}$ where in this case, $\mathbf{b} = (1, 2, 3)^T$. Thus

$$\begin{pmatrix} 1 & 0 & 0 \\ 4 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$

which yields very quickly that $\mathbf{y} = \begin{pmatrix} 1 \\ -2 \\ 2 \end{pmatrix}$. Now you can find $\mathbf{x}$ by solving $U\mathbf{x} = \mathbf{y}$. Thus in this case,

$$\begin{pmatrix} 1 & 2 & 3 & 2 \\ 0 & -5 & -11 & -7 \\ 0 & 0 & 0 & -2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} = \begin{pmatrix} 1 \\ -2 \\ 2 \end{pmatrix}$$

which yields

$$\mathbf{x} = \begin{pmatrix} -\frac{3}{5} + \frac{7}{5}t \\ \frac{9}{5} - \frac{11}{5}t \\ t \\ -1 \end{pmatrix}, t \in \mathbb{R}.$$

Work this out by hand and you will see the advantage of working only with triangular matrices.

It may seem like a trivial thing but it is used because it cuts down on the number of operations involved in finding a solution to a system of equations enough that it makes a difference for large systems.

5.4 The PLU Factorization

As indicated above, some matrices don't have an LU factorization. Here is an example.

$$M = \begin{pmatrix} 1 & 2 & 3 & 2 \\ 1 & 2 & 3 & 0 \\ 4 & 3 & 1 & 1 \end{pmatrix} \tag{5.1}$$

In this case, there is another factorization which is useful called a PLU factorization. Here P is a permutation matrix.

Example 5.4.1 *Find a PLU factorization for the above matrix in 5.1.*

Proceed as before trying to find the row echelon form of the matrix. First add -1 times the first row to the second row and then add -4 times the first to the third. This yields

$$\begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 4 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 2 \\ 0 & 0 & 0 & -2 \\ 0 & -5 & -11 & -7 \end{pmatrix}$$

There is no way to do only row operations involving replacing a row with itself added to a multiple of another row to the second matrix in such a way as to obtain an upper triangular matrix. Therefore, consider M with the bottom two rows switched.

$$M' = \begin{pmatrix} 1 & 2 & 3 & 2 \\ 4 & 3 & 1 & 1 \\ 1 & 2 & 3 & 0 \end{pmatrix}.$$

Now try again with this matrix. First take -1 times the first row and add to the bottom row and then take -4 times the first row and add to the second row. This yields

$$\begin{pmatrix} 1 & 0 & 0 \\ 4 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 2 \\ 0 & -5 & -11 & -7 \\ 0 & 0 & 0 & -2 \end{pmatrix}$$

The second matrix is upper triangular and so the LU factorization of the matrix M' is

$$\begin{pmatrix} 1 & 0 & 0 \\ 4 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 2 \\ 0 & -5 & -11 & -7 \\ 0 & 0 & 0 & -2 \end{pmatrix}.$$

Thus $M' = PM = LU$ where L and U are given above. Therefore, $M = P^2M = PLU$ and so

$$\begin{pmatrix} 1 & 2 & 3 & 2 \\ 1 & 2 & 3 & 0 \\ 4 & 3 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 4 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 2 \\ 0 & -5 & -11 & -7 \\ 0 & 0 & 0 & -2 \end{pmatrix}$$

This process can always be followed and so there always exists a PLU factorization of a given matrix even though there isn't always an LU factorization.

Example 5.4.2 *Use a PLU factorization of* $M \equiv \begin{pmatrix} 1 & 2 & 3 & 2 \\ 1 & 2 & 3 & 0 \\ 4 & 3 & 1 & 1 \end{pmatrix}$ *to solve the system* $M\mathbf{x} = \mathbf{b}$ *where* $\mathbf{b} = (1,2,3)^T$.

Let $U\mathbf{x} = \mathbf{y}$ and consider $PL\mathbf{y} = \mathbf{b}$. In other words, solve,

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 4 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}.$$

Then multiplying both sides by P gives

$$\begin{pmatrix} 1 & 0 & 0 \\ 4 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \\ 2 \end{pmatrix}$$

and so

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}.$$

Now $U\mathbf{x} = \mathbf{y}$ and so it only remains to solve

$$\begin{pmatrix} 1 & 2 & 3 & 2 \\ 0 & -5 & -11 & -7 \\ 0 & 0 & 0 & -2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$$

which yields

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} \frac{1}{5} + \frac{7}{5}t \\ \frac{9}{10} - \frac{11}{5}t \\ t \\ -\frac{1}{2} \end{pmatrix} : t \in \mathbb{R}.$$

5.5 Justification For The Multiplier Method

Why does the multiplier method work for finding an LU factorization? Suppose A is a matrix which has the property that the row reduced echelon form for A may be achieved using only the row operations which involve replacing a row with itself added to a multiple of another row. It is not ever necessary to switch rows. Thus every row which is replaced using this row operation in obtaining the echelon form may be modified by using a row which is above it. Furthermore, in the multiplier method for finding the LU factorization, we zero out the elements below the pivot entry in first column and then the next and so on when scanning from the left. In terms of elementary matrices, this means the row operations used to reduce A to upper triangular form correspond to multiplication on the left by lower triangular matrices having all ones down the main diagonal and the sequence of elementary matrices which row reduces A has the property that in scanning the list of elementary matrices from the right to the left, this list consists of several matrices which involve only changes from the identity in the first column, then several which involve only changes from the identity in the second column and so forth. More precisely, $E_p \cdots E_1 A = U$ where U is upper triangular, each E_i is a lower triangular elementary matrix having all ones down the main diagonal, for some r_i, each of $E_{r_1} \cdots E_1$ differs from the identity only in the first column, each of $E_{r_2} \cdots E_{r_1+1}$ differs from the identity only in the second column and so forth. Therefore, $A = \overbrace{E_1^{-1} \cdots E_{p-1}^{-1} E_p^{-1}}^{\text{Will be } L} U$. You multiply the inverses in the reverse order. Now each of the E_i^{-1} is also lower triangular with 1 down the main diagonal. Therefore their product has this property. Recall also that if E_i equals the identity matrix except for having an a in the j^{th} column somewhere below the main diagonal, E_i^{-1} is obtained by replacing the a in E_i with $-a$, thus explaining why we replace with -1 times the multiplier in computing L. In the case where A is a $3 \times m$ matrix, $E_1^{-1} \cdots E_{p-1}^{-1} E_p^{-1}$ is of the form

$$\begin{pmatrix} 1 & 0 & 0 \\ a & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ b & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & c & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ a & 1 & 0 \\ b & c & 1 \end{pmatrix}.$$

Note that scanning from left to right, the first two in the product involve changes in the identity only in the first column while in the third matrix, the change is only in the second. If the entries in the first column had been zeroed out in a different order, the following would have resulted.

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ b & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ a & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & c & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ a & 1 & 0 \\ b & c & 1 \end{pmatrix}$$

However, it is important to be working from the left to the right, one column at a time.

A similar observation holds in any dimension. Multiplying the elementary matrices which involve a change only in the j^{th} column you obtain A equal to an upper triangular, $n \times m$ matrix U which is multiplied by a sequence of lower triangular matrices on its left which is of the following form, in which the a_{ij} are negatives of multipliers used in row reducing to an upper triangular matrix.

$$\begin{pmatrix} 1 & 0 & \cdots & 0 \\ a_{11} & 1 & & \vdots \\ \vdots & & \ddots & 0 \\ a_{1,n-1} & 0 & \cdots & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 0 & a_{2,n-2} & \cdots & 1 \end{pmatrix} \cdots \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & a_{n,n-1} & 1 \end{pmatrix}$$

From the matrix multiplication, this product equals

$$\begin{pmatrix} 1 & & & \\ a_{11} & 1 & & \\ \vdots & & \ddots & \\ a_{1,n-1} & \cdots & a_{n,n-1} & 1 \end{pmatrix}$$

Notice how the end result of the matrix multiplication made no change in the a_{ij}. It just filled in the empty spaces with the a_{ij} which occurred in one of the matrices in the product. This is why, in computing L, it is sufficient to begin with the left column and work column by column toward the right, replacing entries with the negative of the multiplier used in the row operation which produces a zero in that entry.

5.6 Existence For The PLU Factorization

Here I will consider an invertible $n \times n$ matrix and show that such a matrix always has a PLU factorization. More general matrices could also be considered but this is all I will present.

Let A be such an invertible matrix and consider the first column of A. If $A_{11} \neq 0$, use this to zero out everything below it. The entry A_{11} is called the pivot. Thus in this case there is a lower triangular matrix L_1 which has all ones on the diagonal such that

$$L_1 P_1 A = \begin{pmatrix} * & * \\ \mathbf{0} & A_1 \end{pmatrix} \tag{5.2}$$

Here $P_1 = I$. In case $A_{11} = 0$, let r be such that $A_{r1} \neq 0$ and r is the first entry for which this happens. In this case, let P_1 be the permutation matrix which switches the first row and the r^{th} row. Then as before, there exists a lower triangular matrix L_1 which has all ones on the diagonal such that 5.2 holds in this case also. In the first column, this L_1 has zeros between the first row and the r^{th} row.

Go to A_1. Following the same procedure as above, there exists a lower triangular matrix and permutation matrix L_2', P_2' such that

$$L_2' P_2' A_1 = \begin{pmatrix} * & * \\ \mathbf{0} & A_2 \end{pmatrix}$$

Let

$$L_2 = \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & L_2' \end{pmatrix}, \; P_2 = \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & P_2' \end{pmatrix}$$

Then using block multiplication, Theorem 3.5.2,

$$\begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & L_2' \end{pmatrix} \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & P_2' \end{pmatrix} \begin{pmatrix} * & * \\ \mathbf{0} & A_1 \end{pmatrix} =$$

$$= \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & L_2' \end{pmatrix} \begin{pmatrix} * & * \\ \mathbf{0} & P_2' A_1 \end{pmatrix} = \begin{pmatrix} * & * \\ \mathbf{0} & L_2' P_2' A_1 \end{pmatrix}$$

$$\begin{pmatrix} * & \cdots & * \\ 0 & * & * \\ \mathbf{0} & \mathbf{0} & A_2 \end{pmatrix} = L_2 P_2 L_1 P_1 A$$

and L_2 has all the subdiagonal entries equal to 0 except possibly some nonzero entries in the second column starting with position r_2 where P_2 switches rows r_2 and 2. Continuing this way, it follows there are lower triangular matrices L_j having all ones down the diagonal and permutation matrices P_i which switch only two rows such that

$$L_{n-1} P_{n-1} L_{n-2} P_{n-2} L_{n-3} \cdots L_2 P_2 L_1 P_1 A = U \tag{5.3}$$

where U is upper triangular. The matrix L_j has all zeros below the main diagonal except for the j^{th} column and even in this column it has zeros between position j and r_j where P_j switches rows j and r_j. Of course in the case where no switching is necessary, you could get all nonzero entries below the main diagonal in the j^{th} column for L_j.

The fact that L_j is the identity except for the j^{th} column means that each P_k for $k > j$ almost commutes with L_j. Say P_k switches the k^{th} and the q^{th} rows for $q \geq k > j$. When you place P_k on the right of L_j it just switches the k^{th} and the q^{th} columns and leaves the j^{th} column unchanged. Therefore, the same result as placing P_k on the left of L_j can be obtained by placing P_k on the right of L_j and modifying L_j by switching the k^{th} and the q^{th} entries in the j^{th} column. (Note this could possibly interchange a 0 for something nonzero.) It follows from 5.3 there exists P, the product of permutation matrices, $P = P_{n-1} \cdots P_1$ each of which switches two rows, and L a lower triangular matrix having all ones on the main diagonal, $L = L_{n-1}' \cdots L_2' L_1'$, where the L_j' are obtained as just described by moving a succession of P_k from the left to the right of L_j and modifying the j^{th} column as indicated, such that

$$LPA = U.$$

Then

$$A = P^T L^{-1} U$$

It is customary to write this more simply as

$$A = PLU$$

where L is an upper triangular matrix having all ones on the diagonal and P is a permutation matrix consisting of $P_1 \cdots P_{n-1}$ as described above. This proves the following theorem.

Theorem 5.6.1 *Let A be any invertible $n \times n$ matrix. Then there exists a permutation matrix P and a lower triangular matrix L having all ones on the main diagonal and an upper triangular matrix U such that*

$$A = PLU$$

5.7 The QR Factorization

As pointed out above, the LU factorization is not a mathematically respectable thing because it does not always exist. There is another factorization which does always exist. Much more can be said about it than I will say here. At this time, I will only deal with real matrices and so the inner product will be the usual real dot product.

Definition 5.7.1 *An $n \times n$ real matrix Q is called an orthogonal matrix if*

$$QQ^T = Q^T Q = I.$$

Thus an orthogonal matrix is one whose inverse is equal to its transpose.

First note that if a matrix is orthogonal this says

$$\sum_j Q_{ij}^T Q_{jk} = \sum_j Q_{ji} Q_{jk} = \delta_{ik}$$

Thus

$$\begin{aligned} |Q\mathbf{x}|^2 = \sum_i \left(\sum_j Q_{ij} x_j \right)^2 &= \sum_i \sum_r \sum_s Q_{is} x_s Q_{ir} x_r \\ = \sum_i \sum_r \sum_s Q_{is} Q_{ir} x_s x_r &= \sum_r \sum_s \sum_i Q_{is} Q_{ir} x_s x_r \\ = \sum_r \sum_s \delta_{sr} x_s x_r &= \sum_r x_r^2 = |\mathbf{x}|^2 \end{aligned}$$

This shows that orthogonal transformations preserve distances. You can show that if you have a matrix which does preserve distances, then it must be orthogonal also.

Example 5.7.2 *One of the most important examples of an orthogonal matrix is the so called Householder matrix. You have $\mathbf{v}$ a unit vector and you form the matrix*

$$I - 2\mathbf{v}\mathbf{v}^T$$

This is an orthogonal matrix which is also symmetric. To see this, you use the rules of matrix operations.

$$\begin{aligned} \left(I - 2\mathbf{v}\mathbf{v}^T\right)^T &= I^T - \left(2\mathbf{v}\mathbf{v}^T\right)^T \\ &= I - 2\mathbf{v}\mathbf{v}^T \end{aligned}$$

so it is symmetric. Now to show it is orthogonal,

$$\begin{aligned}\left(I-2\mathbf{v}\mathbf{v}^T\right)\left(I-2\mathbf{v}\mathbf{v}^T\right) &= I-2\mathbf{v}\mathbf{v}^T-2\mathbf{v}\mathbf{v}^T+4\mathbf{v}\mathbf{v}^T\mathbf{v}\mathbf{v}^T \\ &= I-4\mathbf{v}\mathbf{v}^T+4\mathbf{v}\mathbf{v}^T=I\end{aligned}$$

because $\mathbf{v}^T\mathbf{v}=\mathbf{v}\cdot\mathbf{v}=|\mathbf{v}|^2=1$*. Therefore, this is an example of an orthogonal matrix.*

Consider the following problem.

Problem 5.7.3 *Given two vectors* $\mathbf{x},\mathbf{y}$ *such that* $|\mathbf{x}|=|\mathbf{y}|\neq 0$ *but* $\mathbf{x}\neq\mathbf{y}$ *and you want an orthogonal matrix* Q *such that* $Q\mathbf{x}=\mathbf{y}$ *and* $Q\mathbf{y}=\mathbf{x}$*. The thing which works is the Householder matrix*

$$Q\equiv I-2\frac{\mathbf{x}-\mathbf{y}}{|\mathbf{x}-\mathbf{y}|^2}(\mathbf{x}-\mathbf{y})^T$$

Here is why this works.

$$\begin{aligned}Q(\mathbf{x}-\mathbf{y}) &= (\mathbf{x}-\mathbf{y})-2\frac{\mathbf{x}-\mathbf{y}}{|\mathbf{x}-\mathbf{y}|^2}(\mathbf{x}-\mathbf{y})^T(\mathbf{x}-\mathbf{y}) \\ &= (\mathbf{x}-\mathbf{y})-2\frac{\mathbf{x}-\mathbf{y}}{|\mathbf{x}-\mathbf{y}|^2}|\mathbf{x}-\mathbf{y}|^2=\mathbf{y}-\mathbf{x}\end{aligned}$$

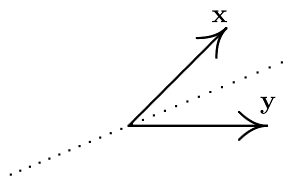
$$\begin{aligned}Q(\mathbf{x}+\mathbf{y}) &= (\mathbf{x}+\mathbf{y})-2\frac{\mathbf{x}-\mathbf{y}}{|\mathbf{x}-\mathbf{y}|^2}(\mathbf{x}-\mathbf{y})^T(\mathbf{x}+\mathbf{y}) \\ &= (\mathbf{x}+\mathbf{y})-2\frac{\mathbf{x}-\mathbf{y}}{|\mathbf{x}-\mathbf{y}|^2}((\mathbf{x}-\mathbf{y})\cdot(\mathbf{x}+\mathbf{y})) \\ &= (\mathbf{x}+\mathbf{y})-2\frac{\mathbf{x}-\mathbf{y}}{|\mathbf{x}-\mathbf{y}|^2}\left(|\mathbf{x}|^2-|\mathbf{y}|^2\right)=\mathbf{x}+\mathbf{y}\end{aligned}$$

Hence

$$\begin{aligned}Q\mathbf{x}+Q\mathbf{y} &= \mathbf{x}+\mathbf{y} \\ Q\mathbf{x}-Q\mathbf{y} &= \mathbf{y}-\mathbf{x}\end{aligned}$$

Adding these equations, $2Q\mathbf{x}=2\mathbf{y}$ *and subtracting them yields* $2Q\mathbf{y}=2\mathbf{x}$*.*

A picture of the geometric significance follows.



The orthogonal matrix Q reflects across the dotted line taking $\mathbf{x}$ to $\mathbf{y}$ and $\mathbf{y}$ to $\mathbf{x}$.

Definition 5.7.4 *Let A be an $m \times n$ matrix. Then a QR factorization of A consists of two matrices, Q orthogonal and R upper triangular (right triangular) having all the entries on the main diagonal nonnegative such that $A = QR$.*

With the solution to this simple problem, here is how to obtain a QR factorization for any matrix A. Let

$$A = (\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n)$$

where the $\mathbf{a}_i$ are the columns. If $\mathbf{a}_1 = \mathbf{0}$, let $Q_1 = I$. If $\mathbf{a}_1 \neq \mathbf{0}$, let

$$\mathbf{b} \equiv \begin{pmatrix} |\mathbf{a}_1| \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

and form the Householder matrix

$$Q_1 \equiv I - 2\frac{(\mathbf{a}_1 - \mathbf{b})}{|\mathbf{a}_1 - \mathbf{b}|^2} (\mathbf{a}_1 - \mathbf{b})^T$$

As in the above problem $Q_1\mathbf{a}_1 = \mathbf{b}$ and so

$$Q_1 A = \begin{pmatrix} |\mathbf{a}_1| & * \\ \mathbf{0} & A_2 \end{pmatrix}$$

where A_2 is a $m-1 \times n-1$ matrix. Now find in the same way as was just done a $m-1 \times m-1$ matrix $\widehat{Q}_2$ such that

$$\widehat{Q}_2 A_2 = \begin{pmatrix} * & * \\ \mathbf{0} & A_3 \end{pmatrix}$$

Let

$$Q_2 \equiv \begin{pmatrix} 1 & 0 \\ \mathbf{0} & \widehat{Q}_2 \end{pmatrix}.$$

Then

$$\begin{aligned} Q_2 Q_1 A &= \begin{pmatrix} 1 & 0 \\ \mathbf{0} & \widehat{Q}_2 \end{pmatrix} \begin{pmatrix} |\mathbf{a}_1| & * \\ \mathbf{0} & A_2 \end{pmatrix} \\ &= \begin{pmatrix} |\mathbf{a}_1| & * & * \\ \vdots & * & * \\ 0 & \mathbf{0} & A_3 \end{pmatrix} \end{aligned}$$

Continuing this way until the result is upper triangular, you get a sequence of orthogonal matrices $Q_p Q_{p-1} \cdots Q_1$ such that

$$Q_p Q_{p-1} \cdots Q_1 A = R \tag{5.4}$$

where R is upper triangular.

Now if Q_1 and Q_2 are orthogonal, then from properties of matrix multiplication,

$$Q_1 Q_2 (Q_1 Q_2)^T = Q_1 Q_2 Q_2^T Q_1^T = Q_1 I Q_1^T = I$$

and similarly

$$(Q_1 Q_2)^T Q_1 Q_2 = I.$$

Thus the product of orthogonal matrices is orthogonal. Also the transpose of an orthogonal matrix is orthogonal directly from the definition. Therefore, from 5.4

$$A = (Q_p Q_{p-1} \cdots Q_1)^T R \equiv QR.$$

This proves the following theorem.

Theorem 5.7.5 *Let A be any real $m \times n$ matrix. Then there exists an orthogonal matrix Q and an upper triangular matrix R having nonnegative entries on the main diagonal such that*

$$A = QR$$

and this factorization can be accomplished in a systematic manner.

►►

5.8 Exercises

1. Find a LU factorization of $\begin{pmatrix} 1 & 2 & 0 \\ 2 & 1 & 3 \\ 1 & 2 & 3 \end{pmatrix}$.

2. Find a LU factorization of $\begin{pmatrix} 1 & 2 & 3 & 2 \\ 1 & 3 & 2 & 1 \\ 5 & 0 & 1 & 3 \end{pmatrix}$.

3. Find a PLU factorization of $\begin{pmatrix} 1 & 2 & 1 \\ 1 & 2 & 2 \\ 2 & 1 & 1 \end{pmatrix}$.

4. Find a PLU factorization of $\begin{pmatrix} 1 & 2 & 1 & 2 & 1 \\ 2 & 4 & 2 & 4 & 1 \\ 1 & 2 & 1 & 3 & 2 \end{pmatrix}$.

5. Find a PLU factorization of $\begin{pmatrix} 1 & 2 & 1 \\ 1 & 2 & 2 \\ 2 & 4 & 1 \\ 3 & 2 & 1 \end{pmatrix}$.

6. Is there only one LU factorization for a given matrix? **Hint:** Consider the equation
$$\begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

7. Here is a matrix and an LU factorization of it.
$$A = \begin{pmatrix} 1 & 2 & 5 & 0 \\ 1 & 1 & 4 & 9 \\ 0 & 1 & 2 & 5 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 5 & 0 \\ 0 & -1 & -1 & 9 \\ 0 & 0 & 1 & 14 \end{pmatrix}$$
Use this factorization to solve the system of equations
$$A\mathbf{x} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$

8. Find a QR factorization for the matrix
$$\begin{pmatrix} 1 & 2 & 1 \\ 3 & -2 & 1 \\ 1 & 0 & 2 \end{pmatrix}$$

9. Find a QR factorization for the matrix
$$\begin{pmatrix} 1 & 2 & 1 & 0 \\ 3 & 0 & 1 & 1 \\ 1 & 0 & 2 & 1 \end{pmatrix}$$

10. If you had a QR factorization, $A = QR$, describe how you could use it to solve the equation $A\mathbf{x} = \mathbf{b}$.

11. If Q is an orthogonal matrix, show the columns are an orthonormal set. That is show that for
$$Q = \begin{pmatrix} \mathbf{q}_1 & \cdots & \mathbf{q}_n \end{pmatrix}$$
it follows that $\mathbf{q}_i \cdot \mathbf{q}_j = \delta_{ij}$. Also show that any orthonormal set of vectors is linearly independent.

12. Show you can't expect uniqueness for QR factorizations. Consider
$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$
and verify this equals
$$\begin{pmatrix} 0 & 1 & 0 \\ \frac{1}{2}\sqrt{2} & 0 & \frac{1}{2}\sqrt{2} \\ \frac{1}{2}\sqrt{2} & 0 & -\frac{1}{2}\sqrt{2} \end{pmatrix} \begin{pmatrix} 0 & 0 & \sqrt{2} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$
and also
$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$
Using Definition 5.7.4, can it be concluded that if A is an invertible matrix it will follow there is only one QR factorization?

13. Suppose $\{\mathbf{a}_1, \cdots, \mathbf{a}_n\}$ are linearly independent vectors in $\mathbb{R}^n$ and let
$$A = \begin{pmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{pmatrix}$$
Form a QR factorization for A.
$$\begin{pmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{pmatrix} = \begin{pmatrix} \mathbf{q}_1 & \cdots & \mathbf{q}_n \end{pmatrix} \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ 0 & r_{22} & \cdots & r_{2n} \\ \vdots & & \ddots & \\ 0 & 0 & \cdots & r_{nn} \end{pmatrix}$$
Show that for each $k \leq n$,
$$\operatorname{span}(\mathbf{a}_1, \cdots, \mathbf{a}_k) = \operatorname{span}(\mathbf{q}_1, \cdots, \mathbf{q}_k)$$
Prove that every subspace of $\mathbb{R}^n$ has an orthonormal basis. The procedure just described is similar to the Gram Schmidt procedure which will be presented later.

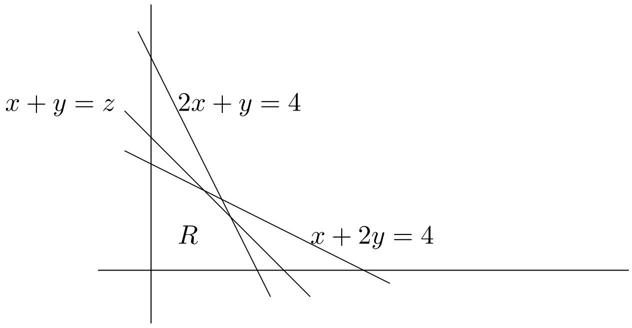
14. Suppose $Q_n R_n$ converges to an orthogonal matrix Q where Q_n is orthogonal and R_n is upper triangular having all positive entries on the diagonal. Show that then Q_n converges to Q and R_n converges to the identity.

Linear Programming

6.1 Simple Geometric Considerations

One of the most important uses of row operations is in solving linear program problems which involve maximizing a linear function subject to inequality constraints determined from linear equations. Here is an example. A certain hamburger store has 9000 hamburger patties to use in one week and a limitless supply of special sauce, lettuce, tomatoes, onions, and buns. They sell two types of hamburgers, the big stack and the basic burger. It has also been determined that the employees cannot prepare more than 9000 of either type in one week. The big stack, popular with the teenagers from the local high school, involves two patties, lots of delicious sauce, condiments galore, and a divider between the two patties. The basic burger, very popular with children, involves only one patty and some pickles and ketchup. Demand for the basic burger is twice what it is for the big stack. What is the maximum number of hamburgers which could be sold in one week given the above limitations?

Let x be the number of basic burgers and y the number of big stacks which could be sold in a week. Thus it is desired to maximize $z = x + y$ subject to the above constraints. The total number of patties is 9000 and so the number of patty used is $x+2y$. This number must satisfy $x+2y \leq 9000$ because there are only 9000 patty available. Because of the limitation on the number the employees can prepare and the demand, it follows $2x + y \leq 9000$. You never sell a negative number of hamburgers and so $x, y \geq 0$. In simpler terms the problem reduces to maximizing $z = x + y$ subject to the two constraints, $x + 2y \leq 9000$ and $2x + y \leq 9000$. This problem is pretty easy to solve geometrically. Consider the following picture in which R labels the region described by the above inequalities and the line $z = x+y$ is shown for a particular value of z.



As you make z larger this line moves away from the origin, always having the same slope and the desired solution would consist of a point in the region, R which makes z as large as possible or equivalently one for which the line is as far as possible from the origin. Clearly this point is the point of intersection of the two lines, $(3000, 3000)$ and so the maximum value of the given function is 6000. Of course this type of procedure is fine for a situation in which there are only two variables but what about a similar problem in which there are very many variables. In reality, this hamburger store makes many more types of burgers than those two and there are many considerations other than demand and available patty. Each will likely give you a constraint which must be considered in order to solve a more realistic problem and the end result will likely be a problem in many dimensions, probably many more than three so your ability to draw a picture will get you nowhere for such a problem. Another method is needed. This method is the topic of this section. I will illustrate with this particular problem. Let $x_1 = x$ and $y = x_2$. Also let x_3 and x_4 be nonnegative variables such that

$$x_1 + 2x_2 + x_3 = 9000,\ 2x_1 + x_2 + x_4 = 9000.$$

To say that x_3 and x_4 are nonnegative is the same as saying $x_1 + 2x_2 \leq 9000$ and $2x_1 + x_2 \leq 9000$ and these variables are called slack variables at this point. They are called this because they "take up the slack". I will discuss these more later. First a general situation is considered.

6.2 The Simplex Tableau

Here is some notation.

Definition 6.2.1 *Let* $\mathbf{x}, \mathbf{y}$ *be vectors in* $\mathbb{R}^q$. *Then* $\mathbf{x} \leq \mathbf{y}$ *means for each* $i, x_i \leq y_i$.

The problem is as follows:

Let A be an $m \times (m+n)$ real matrix of rank m. It is desired to find $\mathbf{x} \in \mathbb{R}^{n+m}$ such that $\mathbf{x}$ satisfies the constraints,

$$\mathbf{x} \geq \mathbf{0},\ A\mathbf{x} = \mathbf{b} \tag{6.1}$$

and out of all such $\mathbf{x}$,

$$z \equiv \sum_{i=1}^{m+n} c_i x_i$$

is as large (or small) as possible. This is usually referred to as maximizing or minimizing z subject to the above constraints. First I will consider the constraints.

Let $A = \begin{pmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_{n+m} \end{pmatrix}$. First you find a vector, $\mathbf{x}^0 \geq \mathbf{0}$, $A\mathbf{x}^0 = \mathbf{b}$ such that n of the components of this vector equal 0. Letting $i_1, \cdots, i_n$ be the positions of $\mathbf{x}^0$ for which $x_i^0 = 0$, suppose also that $\{\mathbf{a}_{j_1}, \cdots, \mathbf{a}_{j_m}\}$ is linearly independent for j_i the other positions of $\mathbf{x}^0$. Geometrically, this means that $\mathbf{x}^0$ is a corner of the feasible region, those $\mathbf{x}$ which satisfy the constraints. This is called a basic feasible solution. Also define

$$\begin{aligned} \mathbf{c}_B &\equiv (c_{j_1}, \cdots, c_{j_m}),\ \mathbf{c}_F \equiv (c_{i_1}, \cdots, c_{i_n}) \\ \mathbf{x}_B &\equiv (x_{j_1}, \cdots, x_{j_m}),\ \mathbf{x}_F \equiv (x_{i_1}, \cdots, x_{i_n}). \end{aligned}$$

and

$$z^0 \equiv z\left(\mathbf{x}^0\right) = \begin{pmatrix} \mathbf{c}_B & \mathbf{c}_F \end{pmatrix} \begin{pmatrix} \mathbf{x}_B^0 \\ \mathbf{x}_F^0 \end{pmatrix} = \mathbf{c}_B \mathbf{x}_B^0$$

since $\mathbf{x}_F^0 = \mathbf{0}$. The variables which are the components of the vector $\mathbf{x}_B$ are called the **basic variables** and the variables which are the entries of $\mathbf{x}_F$ are called the **free variables.** You set $\mathbf{x}_F = \mathbf{0}$. Now $\left(\mathbf{x}^0, z^0\right)^T$ is a solution to

$$\begin{pmatrix} A & \mathbf{0} \\ -\mathbf{c} & 1 \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ z \end{pmatrix} = \begin{pmatrix} \mathbf{b} \\ 0 \end{pmatrix}$$

along with the constraints $\mathbf{x} \geq \mathbf{0}$. Writing the above in augmented matrix form yields

$$\begin{pmatrix} A & \mathbf{0} & \mathbf{b} \\ -\mathbf{c} & 1 & 0 \end{pmatrix} \tag{6.2}$$

Permute the columns and variables on the left if necessary to write the above in the form

$$\begin{pmatrix} B & F & \mathbf{0} \\ -\mathbf{c}_B & -\mathbf{c}_F & 1 \end{pmatrix} \begin{pmatrix} \mathbf{x}_B \\ \mathbf{x}_F \\ z \end{pmatrix} = \begin{pmatrix} \mathbf{b} \\ 0 \end{pmatrix} \tag{6.3}$$

or equivalently in the augmented matrix form keeping track of the variables on the bottom as

$$\begin{pmatrix} B & F & \mathbf{0} & \mathbf{b} \\ -\mathbf{c}_B & -\mathbf{c}_F & 1 & 0 \\ \mathbf{x}_B & \mathbf{x}_F & 0 & 0 \end{pmatrix}. \tag{6.4}$$

Here B pertains to the variables $x_{i_1}, \cdots, x_{j_m}$ and is an $m \times m$ matrix with linearly independent columns, $\{\mathbf{a}_{j_1}, \cdots, \mathbf{a}_{j_m}\}$, and F is an $m \times n$ matrix. Now it is assumed that

$$\begin{pmatrix} B & F \end{pmatrix} \begin{pmatrix} \mathbf{x}_B^0 \\ \mathbf{x}_F^0 \end{pmatrix} = \begin{pmatrix} B & F \end{pmatrix} \begin{pmatrix} \mathbf{x}_B^0 \\ \mathbf{0} \end{pmatrix} = B\mathbf{x}_B^0 = \mathbf{b}$$

and since B is assumed to have rank m, it follows

$$\mathbf{x}_B^0 = B^{-1}\mathbf{b} \geq \mathbf{0}. \tag{6.5}$$

This is very important to observe. $B^{-1}\mathbf{b} \geq \mathbf{0}$! This is by the assumption that $\mathbf{x}^0 \geq \mathbf{0}$.

Do row operations on the top part of the matrix

$$\begin{pmatrix} B & F & \mathbf{0} & \mathbf{b} \\ -\mathbf{c}_B & -\mathbf{c}_F & 1 & 0 \end{pmatrix} \tag{6.6}$$

and obtain its row reduced echelon form. Then after these row operations the above becomes

$$\left(\begin{array}{cccc} I & B^{-1}F & \mathbf{0} & B^{-1}\mathbf{b} \\ -\mathbf{c}_B & -\mathbf{c}_F & 1 & 0 \end{array}\right). \tag{6.7}$$

where $B^{-1}\mathbf{b} \geq 0$. Next do another row operation in order to get a $\mathbf{0}$ where you see a $-\mathbf{c}_B$. Thus

$$\left(\begin{array}{cccc} I & B^{-1}F & \mathbf{0} & B^{-1}\mathbf{b} \\ \mathbf{0} & \mathbf{c}_B B^{-1}F' - \mathbf{c}_F & 1 & \mathbf{c}_B B^{-1}\mathbf{b} \end{array}\right) \tag{6.8}$$

$$= \left(\begin{array}{cccc} I & B^{-1}F & \mathbf{0} & B^{-1}\mathbf{b} \\ \mathbf{0} & \mathbf{c}_B B^{-1}F' - \mathbf{c}_F & 1 & \mathbf{c}_B \mathbf{x}_B^0 \end{array}\right)$$

$$= \left(\begin{array}{cccc} I & B^{-1}F & \mathbf{0} & B^{-1}\mathbf{b} \\ \mathbf{0} & \mathbf{c}_B B^{-1}F - \mathbf{c}_F & 1 & z^0 \end{array}\right) \tag{6.9}$$

The reason there is a z^0 on the bottom right corner is that $\mathbf{x}_F = 0$ and $\left(\mathbf{x}_B^0, \mathbf{x}_F^0, z^0\right)^T$ is a solution of the system of equations represented by the above augmented matrix because it is a solution to the system of equations corresponding to the system of equations represented by 6.6 and row operations leave solution sets unchanged. Note how attractive this is. The z_0 is the value of z at the point $\mathbf{x}^0$. The augmented matrix of 6.9 is called the simplex tableau and it is the beginning point for the simplex algorithm to be described a little later. It is very convenient to express the simplex tableau in the above form in which the variables are possibly permuted in order to have $\left(\begin{array}{c} I \\ \mathbf{0} \end{array}\right)$ on the left side. However, as far as the simplex algorithm is concerned it is not necessary to be permuting the variables in this manner. Starting with 6.9 you could permute the variables and columns to obtain an augmented matrix in which the variables are in their original order. What is really required for the simplex tableau?

It is an augmented $m+1 \times m+n+2$ matrix which represents a system of equations which has the same set of solutions, $(\mathbf{x},z)^T$ as the system whose augmented matrix is

$$\left(\begin{array}{ccc} A & \mathbf{0} & \mathbf{b} \\ -\mathbf{c} & 1 & 0 \end{array}\right)$$

(Possibly the variables for $\mathbf{x}$ are taken in another order.) There are m linearly independent columns in the first $m+n$ columns for which there is only one nonzero entry, a 1 in one of the first m rows, the "simple columns", the other first $m+n$ columns being the "nonsimple columns". As in the above, the variables corresponding to the simple columns are $\mathbf{x}_B$, the basic variables and those corresponding to the nonsimple columns are $\mathbf{x}_F$, the free variables. Also, the top m entries of the last column on the right are nonnegative. This is the description of a simplex tableau.

In a simplex tableau it is easy to spot a basic feasible solution. You can see one quickly by setting the variables, $\mathbf{x}_F$ corresponding to the nonsimple columns equal to zero. Then the other variables, corresponding to the simple columns are each equal to a nonnegative entry in the far right column. Lets call this an "**obvious basic feasible solution**". If a solution is obtained by setting the variables corresponding to the nonsimple columns equal to zero and the variables corresponding to the simple columns equal to zero this will be referred to as an "**obvious**" solution. Lets also call the first $m+n$ entries in the bottom row the "bottom left row". In a simplex tableau, the entry in the bottom right corner gives the value of the variable being maximized or minimized when the obvious basic feasible solution is chosen.

The following is a special case of the general theory presented above and shows how such a special case can be fit into the above framework. The following example is rather typical of the sorts of problems considered. It involves inequality constraints instead of $A\mathbf{x} = \mathbf{b}$. This is handled by adding in "slack variables" as explained below.

The idea is to obtain an augmented matrix for the constraints such that obvious solutions are also feasible. Then there is an algorithm, to be presented later, which takes you from one obvious feasible solution to another until you obtain the maximum.

Example 6.2.2 *Consider $z = x_1 - x_2$ subject to the constraints, $x_1 + 2x_2 \leq 10, x_1 + 2x_2 \geq 2$, and $2x_1 + x_2 \leq 6, x_i \geq 0$. Find a simplex tableau for a problem of the form $\mathbf{x} \geq \mathbf{0}, A\mathbf{x} = \mathbf{b}$ which is equivalent to the above problem.*

You add in slack variables. These are positive variables, one for each of the first three constraints, which change the first three inequalities into equations. Thus the first three inequalities become $x_1 + 2x_2 + x_3 = 10, x_1 + 2x_2 - x_4 = 2$, and $2x_1 + x_2 + x_5 = 6, x_1, x_2, x_3, x_4, x_5 \geq 0$. Now it is necessary to find a basic feasible solution. You mainly need to find a positive solution to the equations,

$$\begin{aligned} x_1 + 2x_2 + x_3 &= 10 \\ x_1 + 2x_2 - x_4 &= 2 \\ 2x_1 + x_2 + x_5 &= 6 \end{aligned}$$

the solution set for the above system is given by

$$x_2 = \frac{2}{3}x_4 - \frac{2}{3} + \frac{1}{3}x_5, x_1 = -\frac{1}{3}x_4 + \frac{10}{3} - \frac{2}{3}x_5, x_3 = -x_4 + 8.$$

An easy way to get a basic feasible solution is to let $x_4 = 8$ and $x_5 = 1$. Then a feasible solution is

$$(x_1, x_2, x_3, x_4, x_5) = (0, 5, 0, 8, 1).$$

It follows $z^0 = -5$ and the matrix 6.2, $\begin{pmatrix} A & \mathbf{0} & \mathbf{b} \\ -\mathbf{c} & 1 & 0 \end{pmatrix}$ with the variables kept track of on the bottom is

$$\begin{pmatrix} 1 & 2 & 1 & 0 & 0 & 0 & 10 \\ 1 & 2 & 0 & -1 & 0 & 0 & 2 \\ 2 & 1 & 0 & 0 & 1 & 0 & 6 \\ -1 & 1 & 0 & 0 & 0 & 1 & 0 \\ x_1 & x_2 & x_3 & x_4 & x_5 & 0 & 0 \end{pmatrix}$$

and the first thing to do is to permute the columns so that the list of variables on the bottom will have x_1 and x_3 at the end.

$$\begin{pmatrix} 2 & 0 & 0 & 1 & 1 & 0 & 10 \\ 2 & -1 & 0 & 1 & 0 & 0 & 2 \\ 1 & 0 & 1 & 2 & 0 & 0 & 6 \\ 1 & 0 & 0 & -1 & 0 & 1 & 0 \\ x_2 & x_4 & x_5 & x_1 & x_3 & 0 & 0 \end{pmatrix}$$

Next, as described above, take the row reduced echelon form of the top three lines of the above matrix. This yields

$$\begin{pmatrix} 1 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 5 \\ 0 & 1 & 0 & 0 & 1 & 0 & 8 \\ 0 & 0 & 1 & \frac{3}{2} & -\frac{1}{2} & 0 & 1 \end{pmatrix}.$$

Now do row operations to

$$\begin{pmatrix} 1 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 5 \\ 0 & 1 & 0 & 0 & 1 & 0 & 8 \\ 0 & 0 & 1 & \frac{3}{2} & -\frac{1}{2} & 0 & 1 \\ 1 & 0 & 0 & -1 & 0 & 1 & 0 \end{pmatrix}$$

to finally obtain

$$\begin{pmatrix} 1 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 5 \\ 0 & 1 & 0 & 0 & 1 & 0 & 8 \\ 0 & 0 & 1 & \frac{3}{2} & -\frac{1}{2} & 0 & 1 \\ 0 & 0 & 0 & -\frac{3}{2} & -\frac{1}{2} & 1 & -5 \end{pmatrix}$$

and this is a simplex tableau. The variables are $x_2, x_4, x_5, x_1, x_3, z$.

It isn't as hard as it may appear from the above. Lets not permute the variables and simply find an acceptable simplex tableau as described above.

Example 6.2.3 *Consider* $z = x_1 - x_2$ *subject to the constraints,* $x_1 + 2x_2 \leq 10, x_1 + 2x_2 \geq 2,$ *and* $2x_1 + x_2 \leq 6, x_i \geq 0$. *Find a simplex tableau.*

Adding in slack variables, an augmented matrix which is descriptive of the constraints is

$$\begin{pmatrix} 1 & 2 & 1 & 0 & 0 & 10 \\ 1 & 2 & 0 & -1 & 0 & 6 \\ 2 & 1 & 0 & 0 & 1 & 6 \end{pmatrix}$$

The obvious solution is not feasible because of that -1 in the fourth column. When you let $x_1, x_2 = 0$, you end up having $x_4 = -6$ which is negative. Consider the second column and select the 2 as a pivot to zero out that which is above and below the 2.

$$\begin{pmatrix} 0 & 0 & 1 & 1 & 0 & 4 \\ \frac{1}{2} & 1 & 0 & -\frac{1}{2} & 0 & 3 \\ \frac{3}{2} & 0 & 0 & \frac{1}{2} & 1 & 3 \end{pmatrix}$$

This one is good. When you let $x_1 = x_4 = 0$, you find that $x_2 = 3, x_3 = 4, x_5 = 3$. The obvious solution is now feasible. You can now assemble the simplex tableau. The first step is to include a column and row for z. This yields

$$\begin{pmatrix} 0 & 0 & 1 & 1 & 0 & 0 & 4 \\ \frac{1}{2} & 1 & 0 & -\frac{1}{2} & 0 & 0 & 3 \\ \frac{3}{2} & 0 & 0 & \frac{1}{2} & 1 & 0 & 3 \\ -1 & 0 & 1 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Now you need to get zeros in the right places so the simple columns will be preserved as simple columns in this larger matrix. This means you need to zero out the 1 in the third column on the bottom. A simplex tableau is now

$$\begin{pmatrix} 0 & 0 & 1 & 1 & 0 & 0 & 4 \\ \frac{1}{2} & 1 & 0 & -\frac{1}{2} & 0 & 0 & 3 \\ \frac{3}{2} & 0 & 0 & \frac{1}{2} & 1 & 0 & 3 \\ -1 & 0 & 0 & -1 & 0 & 1 & -4 \end{pmatrix}.$$

Note it is not the same one obtained earlier. There is no reason a simplex tableau should be unique. In fact, it follows from the above general description that you have one for each basic feasible point of the region determined by the constraints.

6.3 The Simplex Algorithm

6.3.1 Maximums

The simplex algorithm takes you from one basic feasible solution to another while maximizing or minimizing the function you are trying to maximize or minimize. Algebraically, it takes you from one simplex tableau to another in which the lower right corner either increases in the case of maximization or decreases in the case of minimization.

I will continue writing the simplex tableau in such a way that the simple columns having only one entry nonzero are on the left. As explained above, this amounts to permuting the variables. I will do this because it is possible to describe what is going on without onerous notation. However, in the examples, I won't worry so much about it. Thus, from a basic feasible solution, a simplex tableau of the following form has been obtained in which the columns for the basic variables, $\mathbf{x}_B$ are listed first and $\mathbf{b} \geq \mathbf{0}$.

$$\begin{pmatrix} I & F & \mathbf{0} & \mathbf{b} \\ \mathbf{0} & \mathbf{c} & 1 & z^0 \end{pmatrix} \tag{6.10}$$

Let $x_i^0 = b_i$ for $i = 1, \cdots, m$ and $x_i^0 = 0$ for $i > m$. Then $\left(\mathbf{x}^0, z^0\right)$ is a solution to the above system and since $\mathbf{b} \geq \mathbf{0}$, it follows $\left(\mathbf{x}^0, z^0\right)$ is a basic feasible solution.

If $c_i < 0$ for some i, and if $F_{ji} \leq 0$ so that a whole column of $\begin{pmatrix} F \\ \mathbf{c} \end{pmatrix}$ is ≤ 0 with the bottom entry < 0, then letting x_i be the variable corresponding to that column, you could leave all the other entries of $\mathbf{x}_F$ equal to zero but change x_i to be positive. Let the new vector be denoted by $\mathbf{x}'_F$ and letting $\mathbf{x}'_B = \mathbf{b} - F\mathbf{x}'_F$ it follows

$$\begin{aligned} \left(\mathbf{x}'_B\right)_k &= b_k - \sum_j F_{kj} \left(\mathbf{x}_F\right)_j \\ &= b_k - F_{ki} x_i \geq 0 \end{aligned}$$

Now this shows $\left(\mathbf{x}'_B, \mathbf{x}'_F\right)$ is feasible whenever $x_i > 0$ and so you could let x_i become arbitrarily large and positive and conclude there is no maximum for z because

$$z = \left(-c_i\right) x_i + z^0 \tag{6.11}$$

If this happens in a simplex tableau, you can say there is no maximum and stop.

What if $\mathbf{c} \geq \mathbf{0}$? Then $z = z^0 - \mathbf{c}\mathbf{x}_F$ and to satisfy the constraints, you need $\mathbf{x}_F \geq 0$. Therefore, in this case, z^0 is the largest possible value of z and so the maximum has been found. You stop when this occurs. Next I explain what to do if neither of the above stopping conditions hold.

The only case which remains is that some $c_i < 0$ and some $F_{ji} > 0$. You pick a column in $\begin{pmatrix} F \\ \mathbf{c} \end{pmatrix}$ in which $c_i < 0$, usually the one for which c_i is the largest in absolute value. You pick $F_{ji} > 0$ as a pivot element, divide the j^{th} row by F_{ji} and then use to obtain zeros above F_{ji} and below F_{ji}, thus obtaining a new simple column. This row operation also makes exactly one of the other simple columns into a nonsimple column. (In terms of variables, it is said that a free variable becomes a basic variable and a basic variable becomes a free variable.) Now permuting the columns and variables, yields

$$\begin{pmatrix} I & F' & \mathbf{0} & \mathbf{b}' \\ \mathbf{0} & \mathbf{c}' & 1 & z^{0\prime} \end{pmatrix}$$

where $z^{0\prime} \geq z^0$ because $z^{0\prime} = z^0 - c_i \left(\frac{b_j}{F_{ji}}\right)$ and $c_i < 0$. If $\mathbf{b}' \geq 0$, you are in the same position you were at the beginning but now z^0 is larger. Now here is the **important** thing. You don't pick just any F_{ji} when you do these row operations. You **pick the positive one for which the row operation results in $\mathbf{b}' \geq \mathbf{0}$**. Otherwise the obvious basic feasible solution obtained by letting $\mathbf{x}'_F = \mathbf{0}$ will fail to satisfy the constraint that $\mathbf{x} \geq \mathbf{0}$.

How is this done? You need

$$b'_k \equiv b_k - \frac{F_{ki} b_j}{F_{ji}} \geq 0 \tag{6.12}$$

for each $k = 1, \cdots, m$ or equivalently,

$$b_k \geq \frac{F_{ki} b_j}{F_{ji}}. \tag{6.13}$$

Now if $F_{ki} \leq 0$ the above holds. Therefore, you only need to check F_{pi} for $F_{pi} > 0$. The pivot, F_{ji} is the one which makes the quotients of the form

$$\frac{b_p}{F_{pi}}$$

for all positive F_{pi} the smallest. This will work because for $F_{ki} > 0$,

$$\frac{b_p}{F_{pi}} \leq \frac{b_k}{F_{ki}} \Rightarrow b_k \geq \frac{F_{ki} b_p}{F_{pi}}$$

Having gotten a new simplex tableau, you do the same thing to it which was just done and continue. As long as $\mathbf{b} > \mathbf{0}$, so you don't encounter the degenerate case, the values for z associated with setting $\mathbf{x}_F = \mathbf{0}$ keep getting strictly larger every time the process is repeated. You keep going until you find $\mathbf{c} \geq \mathbf{0}$. Then you stop. You are at a maximum. Problems can occur in the process in the so called degenerate case when at some stage of the process some $b_j = 0$. In this case you can cycle through different values for $\mathbf{x}$ with no improvement in z. This case will not be discussed here.

Example 6.3.1 *Maximize* $2x_1 + 3x_2$ *subject to the constraints* $x_1 + x_2 \geq 1, 2x_1 + x_2 \leq 6, x_1 + 2x_2 \leq 6,\ x_1, x_2 \geq 0.$

The constraints are of the form

$$\begin{aligned} x_1 + x_2 - x_3 &= 1 \\ 2x_1 + x_2 + x_4 &= 6 \\ x_1 + 2x_2 + x_5 &= 6 \end{aligned}$$

where the x_3, x_4, x_5 are the slack variables. An augmented matrix for these equations is of the form

$$\begin{pmatrix} 1 & 1 & -1 & 0 & 0 & 1 \\ 2 & 1 & 0 & 1 & 0 & 6 \\ 1 & 2 & 0 & 0 & 1 & 6 \end{pmatrix}$$

Obviously the obvious solution is not feasible. It results in $x_3 < 0$. We need to exchange basic variables. Lets just try something.

$$\begin{pmatrix} 1 & 1 & -1 & 0 & 0 & 1 \\ 0 & -1 & 2 & 1 & 0 & 4 \\ 0 & 1 & 1 & 0 & 1 & 5 \end{pmatrix}$$

Now this one is all right because the obvious solution is feasible. Letting $x_2 = x_3 = 0$, it follows that the obvious solution is feasible. Now we add in the objective function as described above.

$$\begin{pmatrix} 1 & 1 & -1 & 0 & 0 & 0 & 1 \\ 0 & -1 & 2 & 1 & 0 & 0 & 4 \\ 0 & 1 & 1 & 0 & 1 & 0 & 5 \\ -2 & -3 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Then do row operations to leave the simple columns the same. Then

$$\begin{pmatrix} 1 & 1 & -1 & 0 & 0 & 0 & 1 \\ 0 & -1 & 2 & 1 & 0 & 0 & 4 \\ 0 & 1 & 1 & 0 & 1 & 0 & 5 \\ 0 & -1 & -2 & 0 & 0 & 1 & 2 \end{pmatrix}$$

Now there are negative numbers on the bottom row to the left of the 1. Lets pick the first. (It would be more sensible to pick the second.) The ratios to look at are $5/1, 1/1$ so pick for the pivot the 1 in the second column and first row. This will leave the right column above the lower right corner nonnegative. Thus the next tableau is

$$\begin{pmatrix} 1 & 1 & -1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 0 & 5 \\ -1 & 0 & 2 & 0 & 1 & 0 & 4 \\ 1 & 0 & -3 & 0 & 0 & 1 & 3 \end{pmatrix}$$

There is still a negative number there to the left of the 1 in the bottom row. The new ratios are $4/2, 5/1$ so the new pivot is the 2 in the third column. Thus the next tableau is

$$\begin{pmatrix} \frac{1}{2} & 1 & 0 & 0 & \frac{1}{2} & 0 & 3 \\ \frac{3}{2} & 0 & 0 & 1 & -\frac{1}{2} & 0 & 3 \\ -1 & 0 & 2 & 0 & 1 & 0 & 4 \\ -\frac{1}{2} & 0 & 0 & 0 & \frac{3}{2} & 1 & 9 \end{pmatrix}$$

Still, there is a negative number in the bottom row to the left of the 1 so the process does not stop yet. The ratios are $3/(3/2)$ and $3/(1/2)$ and so the new pivot is that $3/2$ in the first column. Thus the new tableau is

$$\begin{pmatrix} 0 & 1 & 0 & -\frac{1}{3} & \frac{2}{3} & 0 & 2 \\ \frac{3}{2} & 0 & 0 & 1 & -\frac{1}{2} & 0 & 3 \\ 0 & 0 & 2 & \frac{2}{3} & \frac{2}{3} & 0 & 6 \\ 0 & 0 & 0 & \frac{1}{3} & \frac{4}{3} & 1 & 10 \end{pmatrix}$$

Now stop. The maximum value is 10. This is an easy enough problem to do geometrically and so you can easily verify that this is the right answer. It occurs when $x_4 = x_5 = 0, x_1 = 2, x_2 = 2, x_3 = 3$.

6.3.2 Minimums

How does it differ if you are finding a minimum? From a basic feasible solution, a simplex tableau of the following form has been obtained in which the simple columns for the basic variables, $\mathbf{x}_B$ are listed first and $\mathbf{b} \geq \mathbf{0}$.

$$\begin{pmatrix} I & F & \mathbf{0} & \mathbf{b} \\ \mathbf{0} & \mathbf{c} & 1 & z^0 \end{pmatrix} \tag{6.14}$$

Let $x_i^0 = b_i$ for $i = 1, \cdots, m$ and $x_i^0 = 0$ for $i > m$. Then $(\mathbf{x}^0, z^0)$ is a solution to the above system and since $\mathbf{b} \geq \mathbf{0}$, it follows $(\mathbf{x}^0, z^0)$ is a basic feasible solution. So far, there is no change.

Suppose first that some $c_i > 0$ and $F_{ji} \leq 0$ for each j. Then let $\mathbf{x}'_F$ consist of changing x_i by making it positive but leaving the other entries of $\mathbf{x}_F$ equal to 0. Then from the bottom row,

$$z = -c_i x_i + z^0$$

and you let $\mathbf{x}'_B = \mathbf{b} - F\mathbf{x}'_F \geq \mathbf{0}$. Thus the constraints continue to hold when x_i is made increasingly positive and it follows from the above equation that there is no minimum for z. You stop when this happens.

Next suppose $\mathbf{c} \leq \mathbf{0}$. Then in this case, $z = z^0 - \mathbf{c}\mathbf{x}_F$ and from the constraints, $\mathbf{x}_F \geq \mathbf{0}$ and so $-\mathbf{c}\mathbf{x}_F \geq 0$ and so z^0 is the minimum value and you stop since this is what you are looking for.

What do you do in the case where some $c_i > 0$ and some $F_{ji} > 0$? In this case, you use the simplex algorithm as in the case of maximums to obtain a new simplex tableau in which $z^{0\prime}$ is smaller. You choose F_{ji} the same way to be the positive entry of the i^{th} column such that $b_p/F_{pi} \geq b_j/F_{ji}$ for all positive entries, F_{pi} and do the same row operations. Now this time,

$$z^{0\prime} = z^0 - c_i \left(\frac{b_j}{F_{ji}} \right) < z^0$$

As in the case of maximums no problem can occur and the process will converge unless you have the degenerate case in which some $b_j = 0$. As in the earlier case, this is most unfortunate when it occurs. You see what happens of course. z^0 does not change and the algorithm just delivers different values of the variables forever with no improvement.

To summarize the geometrical significance of the simplex algorithm, it takes you from one corner of the feasible region to another. You go in one direction to find the maximum and in another to find the minimum. For the maximum you try to get rid of negative entries of $\mathbf{c}$ and for minimums you try to eliminate positive entries of $\mathbf{c}$, where the method of elimination involves the auspicious use of an appropriate pivot element and row operations.

Now return to Example 6.2.2. It will be modified to be a maximization problem.

Example 6.3.2 *Maximize $z = x_1 - x_2$ subject to the constraints,*

$$x_1 + 2x_2 \leq 10, x_1 + 2x_2 \geq 2,$$

and $2x_1 + x_2 \leq 6, x_i \geq 0.$

Recall this is the same as maximizing $z = x_1 - x_2$ subject to

$$\begin{pmatrix} 1 & 2 & 1 & 0 & 0 \\ 1 & 2 & 0 & -1 & 0 \\ 2 & 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} 10 \\ 2 \\ 6 \end{pmatrix}, \mathbf{x} \geq \mathbf{0},$$

the variables, x_3, x_4, x_5 being slack variables. Recall the simplex tableau was

$$\begin{pmatrix} 1 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 5 \\ 0 & 1 & 0 & 0 & 1 & 0 & 8 \\ 0 & 0 & 1 & \frac{3}{2} & -\frac{1}{2} & 0 & 1 \\ 0 & 0 & 0 & -\frac{3}{2} & -\frac{1}{2} & 1 & -5 \end{pmatrix}$$

with the variables ordered as x_2, x_4, x_5, x_1, x_3 and so $\mathbf{x}_B = (x_2, x_4, x_5)$ and

$$\mathbf{x}_F = (x_1, x_3).$$

Apply the simplex algorithm to the fourth column because $-\frac{3}{2} < 0$ and this is the most negative entry in the bottom row. The pivot is $3/2$ because $1/(3/2) = 2/3 < 5/(1/2)$. Dividing this row by $3/2$ and then using this to zero out the other elements in that column, the new simplex tableau is

$$\begin{pmatrix} 1 & 0 & -\frac{1}{3} & 0 & \frac{2}{3} & 0 & \frac{14}{3} \\ 0 & 1 & 0 & 0 & 1 & 0 & 8 \\ 0 & 0 & \frac{2}{3} & 1 & -\frac{1}{3} & 0 & \frac{2}{3} \\ 0 & 0 & 1 & 0 & -1 & 1 & -4 \end{pmatrix}.$$

Now there is still a negative number in the bottom left row. Therefore, the process should be continued. This time the pivot is the $2/3$ in the top of the column. Dividing the top row by $2/3$ and then using this to zero out the entries below it,

$$\begin{pmatrix} \frac{3}{2} & 0 & -\frac{1}{2} & 0 & 1 & 0 & 7 \\ -\frac{3}{2} & 1 & \frac{1}{2} & 0 & 0 & 0 & 1 \\ \frac{1}{2} & 0 & \frac{1}{2} & 1 & 0 & 0 & 3 \\ \frac{3}{2} & 0 & \frac{1}{2} & 0 & 0 & 1 & 3 \end{pmatrix}.$$

Now all the numbers on the bottom left row are nonnegative so the process stops. Now recall the variables and columns were ordered as x_2, x_4, x_5, x_1, x_3. The solution in terms of x_1 and x_2 is $x_2 = 0$ and $x_1 = 3$ and $z = 3$. Note that in the above, I did not worry about permuting the columns to keep those which go with the basic variables on the left.

Here is a bucolic example.

Example 6.3.3 *Consider the following table.*

	F_1	F_2	F_3	F_4
iron	*1*	*2*	*1*	*3*
protein	*5*	*3*	*2*	*1*
folic acid	*1*	*2*	*2*	*1*
copper	*2*	*1*	*1*	*1*
calcium	*1*	*1*	*1*	*1*

This information is available to a pig farmer and F_i denotes a particular feed. The numbers in the table contain the number of units of a particular nutrient contained in one pound of *the given feed. Thus F_2 has 2 units of iron in one pound. Now suppose the cost of each feed in cents per pound is given in the following table.*

F_1	F_2	F_3	F_4
2	*3*	*2*	*3*

A typical pig needs 5 units of iron, 8 of protein, 6 of folic acid, 7 of copper and 4 of calcium. (The units may change from nutrient to nutrient.) How many pounds of each feed per pig should the pig farmer use in order to minimize his cost?

His problem is to minimize $C \equiv 2x_1 + 3x_2 + 2x_3 + 3x_4$ subject to the constraints

$$\begin{aligned} x_1 + 2x_2 + x_3 + 3x_4 &\geq 5, \\ 5x_1 + 3x_2 + 2x_3 + x_4 &\geq 8, \\ x_1 + 2x_2 + 2x_3 + x_4 &\geq 6, \\ 2x_1 + x_2 + x_3 + x_4 &\geq 7, \\ x_1 + x_2 + x_3 + x_4 &\geq 4. \end{aligned}$$

where each $x_i \geq 0$. Add in the slack variables,

$$\begin{aligned} x_1 + 2x_2 + x_3 + 3x_4 - x_5 &= 5 \\ 5x_1 + 3x_2 + 2x_3 + x_4 - x_6 &= 8 \\ x_1 + 2x_2 + 2x_3 + x_4 - x_7 &= 6 \\ 2x_1 + x_2 + x_3 + x_4 - x_8 &= 7 \\ x_1 + x_2 + x_3 + x_4 - x_9 &= 4 \end{aligned}$$

The augmented matrix for this system is

$$\begin{pmatrix} 1 & 2 & 1 & 3 & -1 & 0 & 0 & 0 & 0 & 5 \\ 5 & 3 & 2 & 1 & 0 & -1 & 0 & 0 & 0 & 8 \\ 1 & 2 & 2 & 1 & 0 & 0 & -1 & 0 & 0 & 6 \\ 2 & 1 & 1 & 1 & 0 & 0 & 0 & -1 & 0 & 7 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & -1 & 4 \end{pmatrix}$$

How in the world can you find a basic feasible solution? Remember the simplex algorithm is designed to keep the entries in the right column nonnegative so you use this algorithm a few times till the obvious solution is a basic feasible solution.

Consider the first column. The pivot is the 5. Using the row operations described in the algorithm, you get

$$\begin{pmatrix} 0 & \frac{7}{5} & \frac{3}{5} & \frac{14}{5} & -1 & \frac{1}{5} & 0 & 0 & 0 & \frac{17}{5} \\ 1 & \frac{3}{5} & \frac{2}{5} & \frac{1}{5} & 0 & -\frac{1}{5} & 0 & 0 & 0 & \frac{8}{5} \\ 0 & \frac{7}{5} & \frac{8}{5} & \frac{4}{5} & 0 & \frac{1}{5} & -1 & 0 & 0 & \frac{22}{5} \\ 0 & -\frac{1}{5} & \frac{1}{5} & \frac{3}{5} & 0 & \frac{2}{5} & 0 & -1 & 0 & \frac{19}{5} \\ 0 & \frac{2}{5} & \frac{3}{5} & \frac{4}{5} & 0 & \frac{1}{5} & 0 & 0 & -1 & \frac{12}{5} \end{pmatrix}$$

Now go to the second column. The pivot in this column is the 7/5. This is in a different row than the pivot in the first column so I will use it to zero out everything below it. This will get rid of the zeros in the fifth column and introduce zeros in the second. This yields

$$\begin{pmatrix} 0 & 1 & \frac{3}{7} & 2 & -\frac{5}{7} & \frac{1}{7} & 0 & 0 & 0 & \frac{17}{7} \\ 1 & 0 & \frac{1}{7} & -1 & \frac{3}{7} & -\frac{2}{7} & 0 & 0 & 0 & \frac{1}{7} \\ 0 & 0 & 1 & -2 & 1 & 0 & -1 & 0 & 0 & 1 \\ 0 & 0 & \frac{2}{7} & 1 & -\frac{1}{7} & \frac{3}{7} & 0 & -1 & 0 & \frac{30}{7} \\ 0 & 0 & \frac{3}{7} & 0 & \frac{2}{7} & \frac{1}{7} & 0 & 0 & -1 & \frac{10}{7} \end{pmatrix}$$

Now consider another column, this time the fourth. I will pick this one because it has some negative numbers in it so there are fewer entries to check in looking for a pivot. Unfortunately, the pivot is the top 2 and I don't want to pivot on this because it would destroy the zeros in the second column. Consider the fifth column. It is also not a good choice because the pivot is the second element from the top and this would destroy the zeros in the first column. Consider the sixth column. I can use either of the two bottom entries as the pivot. The matrix is

$$\begin{pmatrix} 0 & 1 & 0 & 2 & -1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & -1 & 1 & 0 & 0 & 0 & -2 & 3 \\ 0 & 0 & 1 & -2 & 1 & 0 & -1 & 0 & 0 & 1 \\ 0 & 0 & -1 & 1 & -1 & 0 & 0 & -1 & 3 & 0 \\ 0 & 0 & 3 & 0 & 2 & 1 & 0 & 0 & -7 & 10 \end{pmatrix}$$

Next consider the third column. The pivot is the 1 in the third row. This yields

$$\begin{pmatrix} 0 & 1 & 0 & 2 & -1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & -2 & 2 \\ 0 & 0 & 1 & -2 & 1 & 0 & -1 & 0 & 0 & 1 \\ 0 & 0 & 0 & -1 & 0 & 0 & -1 & -1 & 3 & 1 \\ 0 & 0 & 0 & 6 & -1 & 1 & 3 & 0 & -7 & 7 \end{pmatrix}.$$

There are still 5 columns which consist entirely of zeros except for one entry. Four of them have that entry equal to 1 but one still has a -1 in it, the -1 being in the fourth column. I need to do the row operations on a nonsimple column which has the pivot in the fourth row. Such a column is the second to the last. The pivot is the 3. The new matrix is

$$\begin{pmatrix} 0 & 1 & 0 & \frac{7}{3} & -1 & 0 & \frac{1}{3} & \frac{1}{3} & 0 & \frac{2}{3} \\ 1 & 0 & 0 & \frac{1}{3} & 0 & 0 & \frac{1}{3} & -\frac{2}{3} & 0 & \frac{8}{3} \\ 0 & 0 & 1 & -2 & 1 & 0 & -1 & 0 & 0 & 1 \\ 0 & 0 & 0 & -\frac{1}{3} & 0 & 0 & -\frac{1}{3} & -\frac{1}{3} & 1 & \frac{1}{3} \\ 0 & 0 & 0 & \frac{11}{3} & -1 & 1 & \frac{2}{3} & -\frac{7}{3} & 0 & \frac{28}{3} \end{pmatrix}. \tag{6.15}$$

Now the obvious basic solution is feasible. You let $x_4 = 0 = x_5 = x_7 = x_8$ and $x_1 = 8/3, x_2 = 2/3, x_3 = 1$, and $x_6 = 28/3$. You don't need to worry too much about this. It is the above matrix which is desired. Now you can assemble the simplex tableau and begin the algorithm. Remember $C \equiv 2x_1 + 3x_2 + 2x_3 + 3x_4$. First add the row and column which deal with C. This yields

$$\begin{pmatrix} 0 & 1 & 0 & \frac{7}{3} & -1 & 0 & \frac{1}{3} & \frac{1}{3} & 0 & 0 & \frac{2}{3} \\ 1 & 0 & 0 & \frac{1}{3} & 0 & 0 & \frac{1}{3} & -\frac{2}{3} & 0 & 0 & \frac{8}{3} \\ 0 & 0 & 1 & -2 & 1 & 0 & -1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & -\frac{1}{3} & 0 & 0 & -\frac{1}{3} & -\frac{1}{3} & 1 & 0 & \frac{1}{3} \\ 0 & 0 & 0 & \frac{11}{3} & -1 & 1 & \frac{2}{3} & -\frac{7}{3} & 0 & 0 & \frac{28}{3} \\ -2 & -3 & -2 & -3 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \tag{6.16}$$

Now you do row operations to keep the simple columns of 6.15 simple in 6.16. Of course you could permute the columns if you wanted but this is not necessary.

This yields the following for a simplex tableau. Now it is a matter of getting rid of the positive entries in the bottom row because you are trying to minimize.

$$\begin{pmatrix} 0 & 1 & 0 & \frac{7}{3} & -1 & 0 & \frac{1}{3} & \frac{1}{3} & 0 & 0 & \frac{2}{3} \\ 1 & 0 & 0 & \frac{1}{3} & 0 & 0 & \frac{1}{3} & -\frac{2}{3} & 0 & 0 & \frac{8}{3} \\ 0 & 0 & 1 & -2 & 1 & 0 & -1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & -\frac{1}{3} & 0 & 0 & -\frac{1}{3} & -\frac{1}{3} & 1 & 0 & \frac{1}{3} \\ 0 & 0 & 0 & \frac{11}{3} & -1 & 1 & \frac{2}{3} & -\frac{7}{3} & 0 & 0 & \frac{28}{3} \\ 0 & 0 & 0 & \frac{2}{3} & -1 & 0 & -\frac{1}{3} & -\frac{1}{3} & 0 & 1 & \frac{28}{3} \end{pmatrix}$$

The most positive of them is the $2/3$ and so I will apply the algorithm to this one first. The pivot is the $7/3$. After doing the row operation the next tableau is

$$\begin{pmatrix} 0 & \frac{3}{7} & 0 & 1 & -\frac{3}{7} & 0 & \frac{1}{7} & \frac{1}{7} & 0 & 0 & \frac{2}{7} \\ 1 & -\frac{1}{7} & 0 & 0 & \frac{1}{7} & 0 & \frac{2}{7} & -\frac{5}{7} & 0 & 0 & \frac{18}{7} \\ 0 & \frac{6}{7} & 1 & 0 & \frac{1}{7} & 0 & -\frac{5}{7} & \frac{2}{7} & 0 & 0 & \frac{11}{7} \\ 0 & \frac{1}{7} & 0 & 0 & -\frac{1}{7} & 0 & -\frac{2}{7} & -\frac{2}{7} & 1 & 0 & \frac{3}{7} \\ 0 & -\frac{11}{7} & 0 & 0 & \frac{4}{7} & 1 & \frac{1}{7} & -\frac{20}{7} & 0 & 0 & \frac{58}{7} \\ 0 & -\frac{2}{7} & 0 & 0 & -\frac{5}{7} & 0 & -\frac{3}{7} & -\frac{3}{7} & 0 & 1 & \frac{64}{7} \end{pmatrix}$$

and you see that all the entries are negative and so the minimum is $64/7$ and it occurs when $x_1 = 18/7, x_2 = 0, x_3 = 11/7, x_4 = 2/7$.

There is no maximum for the above problem. However, I will pretend I don't know this and attempt to use the simplex algorithm. You set up the simiplex tableau the same way. Recall it is

$$\begin{pmatrix} 0 & 1 & 0 & \frac{7}{3} & -1 & 0 & \frac{1}{3} & \frac{1}{3} & 0 & 0 & \frac{2}{3} \\ 1 & 0 & 0 & \frac{1}{3} & 0 & 0 & \frac{1}{3} & -\frac{2}{3} & 0 & 0 & \frac{8}{3} \\ 0 & 0 & 1 & -2 & 1 & 0 & -1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & -\frac{1}{3} & 0 & 0 & -\frac{1}{3} & -\frac{1}{3} & 1 & 0 & \frac{1}{3} \\ 0 & 0 & 0 & \frac{11}{3} & -1 & 1 & \frac{2}{3} & -\frac{7}{3} & 0 & 0 & \frac{28}{3} \\ 0 & 0 & 0 & \frac{2}{3} & -1 & 0 & -\frac{1}{3} & -\frac{1}{3} & 0 & 1 & \frac{28}{3} \end{pmatrix}$$

Now to maximize, you try to get rid of the negative entries in the bottom left row. The most negative entry is the -1 in the fifth column. The pivot is the 1 in the third row of this column. The new tableau is

$$\begin{pmatrix} 0 & 1 & 1 & \frac{1}{3} & 0 & 0 & -\frac{2}{3} & \frac{1}{3} & 0 & 0 & \frac{5}{3} \\ 1 & 0 & 0 & \frac{1}{3} & 0 & 0 & \frac{1}{3} & -\frac{2}{3} & 0 & 0 & \frac{8}{3} \\ 0 & 0 & 1 & -2 & 1 & 0 & -1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & -\frac{1}{3} & 0 & 0 & -\frac{1}{3} & -\frac{1}{3} & 1 & 0 & \frac{1}{3} \\ 0 & 0 & 1 & \frac{5}{3} & 0 & 1 & -\frac{1}{3} & -\frac{7}{3} & 0 & 0 & \frac{31}{3} \\ 0 & 0 & 1 & -\frac{4}{3} & 0 & 0 & -\frac{4}{3} & -\frac{1}{3} & 0 & 1 & \frac{31}{3} \end{pmatrix}.$$

Consider the fourth column. The pivot is the top 1/3. The new tableau is

$$\begin{pmatrix} 0 & 3 & 3 & 1 & 0 & 0 & -2 & 1 & 0 & 0 & 5 \\ 1 & -1 & -1 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 1 \\ 0 & 6 & 7 & 0 & 1 & 0 & -5 & 2 & 0 & 0 & 11 \\ 0 & 1 & 1 & 0 & 0 & 0 & -1 & 0 & 1 & 0 & 2 \\ 0 & -5 & -4 & 0 & 0 & 1 & 3 & -4 & 0 & 0 & 2 \\ 0 & 4 & 5 & 0 & 0 & 0 & -4 & 1 & 0 & 1 & 17 \end{pmatrix}$$

There is still a negative in the bottom, the -4. The pivot in that column is the 3. The algorithm yields

$$\begin{pmatrix} 0 & -\frac{1}{3} & \frac{1}{3} & 1 & 0 & \frac{2}{3} & 0 & -\frac{5}{3} & 0 & 0 & \frac{19}{3} \\ 1 & \frac{2}{3} & \frac{1}{3} & 0 & 0 & -\frac{1}{3} & 0 & \frac{1}{3} & 0 & 0 & \frac{1}{3} \\ 0 & -\frac{7}{3} & \frac{1}{3} & 0 & 1 & \frac{5}{3} & 0 & -\frac{14}{3} & 0 & 0 & \frac{43}{3} \\ 0 & -\frac{2}{3} & -\frac{1}{3} & 0 & 0 & \frac{1}{3} & 0 & -\frac{4}{3} & 1 & 0 & \frac{8}{3} \\ 0 & -\frac{5}{3} & -\frac{4}{3} & 0 & 0 & \frac{1}{3} & 1 & -\frac{4}{3} & 0 & 0 & \frac{2}{3} \\ 0 & -\frac{8}{3} & -\frac{1}{3} & 0 & 0 & \frac{4}{3} & 0 & -\frac{13}{3} & 0 & 1 & \frac{59}{3} \end{pmatrix}$$

Note how z keeps getting larger. Consider the column having the $-13/3$ in it. The pivot is the single positive entry, $1/3$. The next tableau is

$$\begin{pmatrix} 5 & 3 & 2 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 8 \\ 3 & 2 & 1 & 0 & 0 & -1 & 0 & 1 & 0 & 0 & 1 \\ 14 & 7 & 5 & 0 & 1 & -3 & 0 & 0 & 0 & 0 & 19 \\ 4 & 2 & 1 & 0 & 0 & -1 & 0 & 0 & 1 & 0 & 4 \\ 4 & 1 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 2 \\ 13 & 6 & 4 & 0 & 0 & -3 & 0 & 0 & 0 & 1 & 24 \end{pmatrix}.$$

There is a column consisting of all negative entries. There is therefore, no maximum. Note also how there is no way to pick the pivot in that column.

Example 6.3.4 *Minimize $z = x_1 - 3x_2 + x_3$ subject to the constraints $x_1 + x_2 + x_3 \leq 10, x_1 + x_2 + x_3 \geq 2$, $x_1 + x_2 + 3x_3 \leq 8$ and $x_1 + 2x_2 + x_3 \leq 7$ with all variables nonnegative.*

There exists an answer because the region defined by the constraints is closed and bounded. Adding in slack variables you get the following augmented matrix corresponding to the constraints.

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 10 \\ 1 & 1 & 1 & 0 & -1 & 0 & 0 & 2 \\ 1 & 1 & 3 & 0 & 0 & 1 & 0 & 8 \\ 1 & 2 & 1 & 0 & 0 & 0 & 1 & 7 \end{pmatrix}$$

Of course there is a problem with the obvious solution obtained by setting to zero all variables corresponding to a nonsimple column because of the simple column which has the -1 in it. Therefore, I will use the simplex algorithm to make this column non simple. The third column has the 1 in the second row as the pivot so I will use this column. This yields

$$\begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 0 & 0 & 8 \\ 1 & 1 & 1 & 0 & -1 & 0 & 0 & 2 \\ -2 & -2 & 0 & 0 & 3 & 1 & 0 & 2 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 & 5 \end{pmatrix} \tag{6.17}$$

and the obvious solution is feasible. Now it is time to assemble the simplex tableau. First add in the bottom row and second to last column corresponding to the equation for z. This yields

$$\begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 8 \\ 1 & 1 & 1 & 0 & -1 & 0 & 0 & 0 & 2 \\ -2 & -2 & 0 & 0 & 3 & 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 5 \\ -1 & 3 & -1 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Next you need to zero out the entries in the bottom row which are below one of the simple columns in 6.17. This yields the simplex tableau

$$\begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 8 \\ 1 & 1 & 1 & 0 & -1 & 0 & 0 & 0 & 2 \\ -2 & -2 & 0 & 0 & 3 & 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 5 \\ 0 & 4 & 0 & 0 & -1 & 0 & 0 & 1 & 2 \end{pmatrix}.$$

The desire is to minimize this so you need to get rid of the positive entries in the left bottom row. There is only one such entry, the 4. In that column the pivot is the 1 in the second row of this column. Thus the next tableau is

$$\begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 8 \\ 1 & 1 & 1 & 0 & -1 & 0 & 0 & 0 & 2 \\ 0 & 0 & 2 & 0 & 1 & 1 & 0 & 0 & 6 \\ -1 & 0 & -1 & 0 & 2 & 0 & 1 & 0 & 3 \\ -4 & 0 & -4 & 0 & 3 & 0 & 0 & 1 & -6 \end{pmatrix}$$

There is still a positive number there, the 3. The pivot in this column is the 2. Apply the algorithm again. This yields

$$\begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} & 1 & 0 & 0 & -\frac{1}{2} & 0 & \frac{13}{2} \\ \frac{1}{2} & 1 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{7}{2} \\ \frac{1}{2} & 0 & \frac{5}{2} & 0 & 0 & 1 & -\frac{1}{2} & 0 & \frac{9}{2} \\ -\frac{1}{2} & 0 & -\frac{1}{2} & 0 & 1 & 0 & \frac{1}{2} & 0 & \frac{3}{2} \\ -\frac{5}{2} & 0 & -\frac{5}{2} & 0 & 0 & 0 & -\frac{3}{2} & 1 & -\frac{21}{2} \end{pmatrix}.$$

Now all the entries in the left bottom row are nonpositive so the process has stopped. The minimum is $-21/2$. It occurs when $x_1 = 0$, $x_2 = 7/2, x_3 = 0$.

Now consider the same problem but change the word, minimize to the word, maximize.

Example 6.3.5 *Maximize $z = x_1 - 3x_2 + x_3$ subject to the constraints $x_1 + x_2 + x_3 \leq 10, x_1 + x_2 + x_3 \geq 2$, $x_1 + x_2 + 3x_3 \leq 8$ and $x_1 + 2x_2 + x_3 \leq 7$ with all variables nonnegative.*

The first part of it is the same. You wind up with the same simplex tableau,

$$\begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 8 \\ 1 & 1 & 1 & 0 & -1 & 0 & 0 & 0 & 2 \\ -2 & -2 & 0 & 0 & 3 & 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 5 \\ 0 & 4 & 0 & 0 & -1 & 0 & 0 & 1 & 2 \end{pmatrix}$$

but this time, you apply the algorithm to get rid of the negative entries in the left bottom row. There is a -1. Use this column. The pivot is the 3. The next tableau is

$$\begin{pmatrix} \frac{2}{3} & \frac{2}{3} & 0 & 1 & 0 & -\frac{1}{3} & 0 & 0 & \frac{22}{3} \\ \frac{1}{3} & \frac{1}{3} & 1 & 0 & 0 & \frac{1}{3} & 0 & 0 & \frac{8}{3} \\ -\frac{2}{3} & -\frac{2}{3} & 0 & 0 & 1 & \frac{1}{3} & 0 & 0 & \frac{2}{3} \\ \frac{2}{3} & \frac{5}{3} & 0 & 0 & 0 & -\frac{1}{3} & 1 & 0 & \frac{13}{3} \\ -\frac{2}{3} & \frac{10}{3} & 0 & 0 & 0 & \frac{1}{3} & 0 & 1 & \frac{8}{3} \end{pmatrix}$$

There is still a negative entry, the $-2/3$. This will be the new pivot column. The pivot is the $2/3$ on the fourth row. This yields

$$\begin{pmatrix} 0 & -1 & 0 & 1 & 0 & 0 & -1 & 0 & 3 \\ 0 & -\frac{1}{2} & 1 & 0 & 0 & \frac{1}{2} & -\frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 5 \\ 1 & \frac{5}{2} & 0 & 0 & 0 & -\frac{1}{2} & \frac{3}{2} & 0 & \frac{13}{2} \\ 0 & 5 & 0 & 0 & 0 & 0 & 1 & 1 & 7 \end{pmatrix}$$

and the process stops. The maximum for z is 7 and it occurs when $x_1 = 13/2, x_2 = 0, x_3 = 1/2$.

6.4 Finding A Basic Feasible Solution

By now it should be fairly clear that finding a basic feasible solution can create considerable difficulty. Indeed, given a system of linear inequalities along with the requirement that each variable be nonnegative, do there even exist points satisfying all these inequalities? If you have many variables, you can't answer this by drawing a picture. Is there some other way to do this which is more systematic than what was presented above? The answer is yes. It is called the method of artificial variables. I will illustrate this method with an example.

Example 6.4.1 *Find a basic feasible solution to the system* $2x_1+x_2-x_3 \geq 3, x_1+x_2+x_3 \geq 2, x_1 + x_2 + x_3 \leq 7$ *and* $\mathbf{x} \geq \mathbf{0}$.

If you write the appropriate augmented matrix with the slack variables,

$$\begin{pmatrix} 2 & 1 & -1 & -1 & 0 & 0 & 3 \\ 1 & 1 & 1 & 0 & -1 & 0 & 2 \\ 1 & 1 & 1 & 0 & 0 & 1 & 7 \end{pmatrix} \tag{6.18}$$

The obvious solution is not feasible. This is why it would be hard to get started with the simplex method. What is the problem? It is those -1 entries in the fourth and fifth columns. To get around this, you add in artificial variables to get an augmented matrix of the form

$$\begin{pmatrix} 2 & 1 & -1 & -1 & 0 & 0 & 1 & 0 & 3 \\ 1 & 1 & 1 & 0 & -1 & 0 & 0 & 1 & 2 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 7 \end{pmatrix} \tag{6.19}$$

Thus the variables are $x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8$. Suppose you can find a feasible solution to the system of equations represented by the above augmented matrix. Thus all variables are nonnegative. Suppose also that it can be done in such a way that x_8 and x_7 happen to be 0. Then it will follow that $x_1, \cdots, x_6$ is a feasible solution for 6.18. Conversely, if you can find a feasible solution for 6.18, then letting x_7 and x_8 both equal zero, you have obtained a feasible solution to 6.19. Since all variables are nonnegative, x_7 and x_8 both equalling zero is equivalent to saying the minimum of $z = x_7 + x_8$ subject to the constraints represented by the above augmented matrix equals zero. This has proved the following simple observation.

Observation 6.4.2 *There exists a feasible solution to the constraints represented by the augmented matrix of 6.18 and* $\mathbf{x} \geq \mathbf{0}$ *if and only if the minimum of* $x_7 + x_8$ *subject to the constraints of 6.19 and* $\mathbf{x} \geq \mathbf{0}$ *exists and equals* 0.

Of course a similar observation would hold in other similar situations. Now the point of all this is that it is trivial to see a feasible solution to 6.19, namely $x_6 = 7, x_7 = 3, x_8 = 2$ and all the other variables may be set to equal zero. Therefore, it is easy to find an initial simplex tableau for the minimization problem just described. First add the column and row for z

$$\begin{pmatrix} 2 & 1 & -1 & -1 & 0 & 0 & 1 & 0 & 0 & 3 \\ 1 & 1 & 1 & 0 & -1 & 0 & 0 & 1 & 0 & 2 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 7 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & 1 & 0 \end{pmatrix}$$

Next it is necessary to make the last two columns on the bottom left row into simple columns. Performing the row operation, this yields an initial simplex tableau,

$$\begin{pmatrix} 2 & 1 & -1 & -1 & 0 & 0 & 1 & 0 & 0 & 3 \\ 1 & 1 & 1 & 0 & -1 & 0 & 0 & 1 & 0 & 2 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 7 \\ 3 & 2 & 0 & -1 & -1 & 0 & 0 & 0 & 1 & 5 \end{pmatrix}$$

Now the algorithm involves getting rid of the positive entries on the left bottom row. Begin with the first column. The pivot is the 2. An application of the simplex algorithm yields the new tableau

$$\begin{pmatrix} 1 & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 & 0 & \frac{3}{2} \\ 0 & \frac{1}{2} & \frac{3}{2} & \frac{1}{2} & -1 & 0 & -\frac{1}{2} & 1 & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{3}{2} & \frac{1}{2} & 0 & 1 & -\frac{1}{2} & 0 & 0 & \frac{11}{2} \\ 0 & \frac{1}{2} & \frac{3}{2} & \frac{1}{2} & -1 & 0 & -\frac{3}{2} & 0 & 1 & \frac{1}{2} \end{pmatrix}$$

Now go to the third column. The pivot is the 3/2 in the second row. An application of the simplex algorithm yields

$$\begin{pmatrix} 1 & \frac{2}{3} & 0 & -\frac{1}{3} & -\frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} & 0 & \frac{5}{3} \\ 0 & \frac{1}{3} & 1 & \frac{1}{3} & -\frac{2}{3} & 0 & -\frac{1}{3} & \frac{2}{3} & 0 & \frac{1}{3} \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & -1 & 0 & 5 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & 1 & 0 \end{pmatrix} \tag{6.20}$$

and you see there are only nonpositive numbers on the bottom left column so the process stops and yields 0 for the minimum of $z = x_7 + x_8$. As for the other variables, $x_1 = 5/3, x_2 = 0, x_3 = 1/3, x_4 = 0, x_5 = 0, x_6 = 5$. Now as explained in the above observation, this is a basic feasible solution for the original system 6.18.

Now consider a maximization problem associated with the above constraints.

Example 6.4.3 *Maximize* $x_1 - x_2 + 2x_3$ *subject to the constraints,* $2x_1 + x_2 - x_3 \geq 3, x_1 + x_2 + x_3 \geq 2, x_1 + x_2 + x_3 \leq 7$ *and* $\mathbf{x} \geq \mathbf{0}$.

From 6.20 you can immediately assemble an initial simplex tableau. You begin with the first 6 columns and top 3 rows in 6.20. Then add in the column and row for z. This yields

$$\begin{pmatrix} 1 & \frac{2}{3} & 0 & -\frac{1}{3} & -\frac{1}{3} & 0 & 0 & \frac{5}{3} \\ 0 & \frac{1}{3} & 1 & \frac{1}{3} & -\frac{2}{3} & 0 & 0 & \frac{1}{3} \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 5 \\ -1 & 1 & -2 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

and you first do row operations to make the first and third columns simple columns. Thus the next simplex tableau is

$$\begin{pmatrix} 1 & \frac{2}{3} & 0 & -\frac{1}{3} & -\frac{1}{3} & 0 & 0 & \frac{5}{3} \\ 0 & \frac{1}{3} & 1 & \frac{1}{3} & -\frac{2}{3} & 0 & 0 & \frac{1}{3} \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 5 \\ 0 & \frac{7}{3} & 0 & \frac{1}{3} & -\frac{5}{3} & 0 & 1 & \frac{7}{3} \end{pmatrix}$$

You are trying to get rid of negative entries in the bottom left row. There is only one, the $-5/3$. The pivot is the 1. The next simplex tableau is then

$$\begin{pmatrix} 1 & \frac{2}{3} & 0 & -\frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{10}{3} \\ 0 & \frac{1}{3} & 1 & \frac{1}{3} & 0 & \frac{2}{3} & 0 & \frac{11}{3} \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 5 \\ 0 & \frac{7}{3} & 0 & \frac{1}{3} & 0 & \frac{5}{3} & 1 & \frac{32}{3} \end{pmatrix}$$

and so the maximum value of z is $32/3$ and it occurs when $x_1 = 10/3, x_2 = 0$ and $x_3 = 11/3$.

6.5 Duality

You can solve minimization problems by solving maximization problems. You can also go the other direction and solve maximization problems by minimization problems. Sometimes this makes things much easier. To be more specific, the two problems to be considered are
$A.)$ Minimize $z = \mathbf{cx}$ subject to $\mathbf{x} \geq \mathbf{0}$ and $A\mathbf{x} \geq \mathbf{b}$ and
$B.)$ Maximize $w = \mathbf{yb}$ such that $\mathbf{y} \geq \mathbf{0}$ and $\mathbf{y}A \leq \mathbf{c}$,

$$\left(\text{equivalently } A^T\mathbf{y}^T \geq \mathbf{c}^T \text{ and } w = \mathbf{b}^T\mathbf{y}^T\right).$$

In these problems it is assumed A is an $m \times p$ matrix.

I will show how a solution of the first yields a solution of the second and then show how a solution of the second yields a solution of the first. The problems, $A.)$ and $B.)$ are called dual problems.

Lemma 6.5.1 *Let $\mathbf{x}$ be a solution of the inequalities of $A.)$ and let $\mathbf{y}$ be a solution of the inequalities of $B.)$. Then*

$$\mathbf{cx} \geq \mathbf{yb}.$$

and if equality holds in the above, then $\mathbf{x}$ is the solution to $A.)$ and $\mathbf{y}$ is a solution to $B.)$.

Proof: This follows immediately. Since $\mathbf{c} \geq \mathbf{y}A, \mathbf{cx} \geq \mathbf{y}A\mathbf{x} \geq \mathbf{yb}$.

It follows from this lemma that if $\mathbf{y}$ satisfies the inequalities of $B.)$ and $\mathbf{x}$ satisfies the inequalities of $A.)$ then if equality holds in the above lemma, it must be that $\mathbf{x}$ is a solution of $A.)$ and $\mathbf{y}$ is a solution of $B.)$. ■

Now recall that to solve either of these problems using the simplex method, you first add in slack variables. Denote by $\mathbf{x}'$ and $\mathbf{y}'$ the enlarged list of variables. Thus $\mathbf{x}'$ has at least m entries and so does $\mathbf{y}'$ and the inequalities involving A were replaced by equalities whose augmented matrices were of the form

$$\begin{pmatrix} A & -I & \mathbf{b} \end{pmatrix}, \text{ and } \begin{pmatrix} A^T & I & \mathbf{c}^T \end{pmatrix}$$

Then you included the row and column for z and w to obtain

$$\begin{pmatrix} A & -I & \mathbf{0} & \mathbf{b} \\ -\mathbf{c} & \mathbf{0} & 1 & 0 \end{pmatrix} \text{ and } \begin{pmatrix} A^T & I & \mathbf{0} & \mathbf{c}^T \\ -\mathbf{b}^T & \mathbf{0} & 1 & 0 \end{pmatrix}. \tag{6.21}$$

Then the problems have basic feasible solutions if it is possible to permute the first $p + m$ columns in the above two matrices and obtain matrices of the form

$$\begin{pmatrix} B & F & \mathbf{0} & \mathbf{b} \\ -\mathbf{c}_B & -\mathbf{c}_F & 1 & 0 \end{pmatrix} \text{ and } \begin{pmatrix} B_1 & F_1 & \mathbf{0} & \mathbf{c}^T \\ -\mathbf{b}_{B_1}^T & -\mathbf{b}_{F_1}^T & 1 & 0 \end{pmatrix} \tag{6.22}$$

where B, B_1 are invertible $m \times m$ and $p \times p$ matrices and denoting the variables associated with these columns by $\mathbf{x}_B, \mathbf{y}_B$ and those variables associated with F or F_1 by $\mathbf{x}_F$ and $\mathbf{y}_F$, it follows that letting $B\mathbf{x}_B = \mathbf{b}$ and $\mathbf{x}_F = \mathbf{0}$, the resulting vector, $\mathbf{x}'$ is a solution to $\mathbf{x}' \geq \mathbf{0}$ and $\begin{pmatrix} A & -I \end{pmatrix} \mathbf{x}' = \mathbf{b}$ with similar constraints holding for $\mathbf{y}'$. In other words, it is possible to obtain simplex tableaus,

$$\begin{pmatrix} I & B^{-1}F & \mathbf{0} & B^{-1}\mathbf{b} \\ \mathbf{0} & \mathbf{c}_B B^{-1}F - \mathbf{c}_F & 1 & \mathbf{c}_B B^{-1}\mathbf{b} \end{pmatrix}, \begin{pmatrix} I & B_1^{-1}F_1 & \mathbf{0} & B_1^{-1}\mathbf{c}^T \\ \mathbf{0} & \mathbf{b}_{B_1}^T B_1^{-1}F - \mathbf{b}_{F_1}^T & 1 & \mathbf{b}_{B_1}^T B_1^{-1}\mathbf{c}^T \end{pmatrix} \tag{6.23}$$

Similar considerations apply to the second problem. Thus as just described, a basic feasible solution is one which determines a simplex tableau like the above in which you get a feasible solution by setting all but the first m variables equal to zero. The simplex algorithm takes you from one basic feasible solution to another till eventually, if there is no degeneracy, you obtain a basic feasible solution which yields the solution of the problem of interest.

Theorem 6.5.2 *Suppose there exists a solution* $\mathbf{x}$ *to* $A.)$ *where* $\mathbf{x}$ *is a basic feasible solution of the inequalities of* $A.)$*. Then there exists a solution* $\mathbf{y}$ *to* $B.)$ *and* $\mathbf{cx} = \mathbf{by}$*. It is also possible to find* $\mathbf{y}$ *from* $\mathbf{x}$ *using a simple formula.*

Proof: Since the solution to $A.)$ is basic and feasible, there exists a simplex tableau like 6.23 such that $\mathbf{x}'$ can be split into $\mathbf{x}_B$ and $\mathbf{x}_F$ such that $\mathbf{x}_F = 0$ and $\mathbf{x}_B = B^{-1}\mathbf{b}$. Now since it is a minimizer, it follows $\mathbf{c}_B B^{-1}F - \mathbf{c}_F \leq \mathbf{0}$ and the minimum value for $\mathbf{cx}$ is $\mathbf{c}_B B^{-1}\mathbf{b}$. Stating this again, $\mathbf{cx} = \mathbf{c}_B B^{-1}\mathbf{b}$. Is it possible you can take $\mathbf{y} = \mathbf{c}_B B^{-1}$? From Lemma 6.5.1 this will be so if $\mathbf{c}_B B^{-1}$ solves the constraints of problem $B.)$. Is $\mathbf{c}_B B^{-1} \geq 0$? Is $\mathbf{c}_B B^{-1}A \leq \mathbf{c}$? These two conditions are satisfied if and only if $\mathbf{c}_B B^{-1} \begin{pmatrix} A & -I \end{pmatrix} \leq \begin{pmatrix} \mathbf{c} & \mathbf{0} \end{pmatrix}$. Referring to the process of permuting the columns of the first augmented matrix of 6.21 to get 6.22 and doing the same permutations on the columns of $\begin{pmatrix} A & -I \end{pmatrix}$ and $\begin{pmatrix} \mathbf{c} & \mathbf{0} \end{pmatrix}$, the desired inequality holds if and only if $\mathbf{c}_B B^{-1} \begin{pmatrix} B & F \end{pmatrix} \leq \begin{pmatrix} \mathbf{c}_B & \mathbf{c}_F \end{pmatrix}$ which is equivalent to saying $\begin{pmatrix} \mathbf{c}_B & \mathbf{c}_B B^{-1}F \end{pmatrix} \leq \begin{pmatrix} \mathbf{c}_B & \mathbf{c}_F \end{pmatrix}$ and this is true because $\mathbf{c}_B B^{-1}F - \mathbf{c}_F \leq \mathbf{0}$ due to the assumption that $\mathbf{x}$ is a minimizer. The simple formula is just $\mathbf{y} = \mathbf{c}_B B^{-1}$. ■

The proof of the following corollary is similar.

Corollary 6.5.3 *Suppose there exists a solution,* $\mathbf{y}$ *to* $B.)$ *where* $\mathbf{y}$ *is a basic feasible solution of the inequalities of* $B.)$*. Then there exists a solution,* $\mathbf{x}$ *to* $A.)$ *and* $\mathbf{cx} = \mathbf{by}$*. It is also possible to find* $\mathbf{x}$ *from* $\mathbf{y}$ *using a simple formula. In this case, and referring to 6.23, the simple formula is* $\mathbf{x} = B_1^{-T}\mathbf{b}_{B_1}$.

As an example, consider the pig farmers problem. The main difficulty in this problem was finding an initial simplex tableau. Now consider the following example and marvel at how all the difficulties disappear.

Example 6.5.4 *minimize* $C \equiv 2x_1 + 3x_2 + 2x_3 + 3x_4$ *subject to the constraints*

$$\begin{aligned} x_1 + 2x_2 + x_3 + 3x_4 &\geq 5, \\ 5x_1 + 3x_2 + 2x_3 + x_4 &\geq 8, \\ x_1 + 2x_2 + 2x_3 + x_4 &\geq 6, \\ 2x_1 + x_2 + x_3 + x_4 &\geq 7, \\ x_1 + x_2 + x_3 + x_4 &\geq 4. \end{aligned}$$

where each $x_i \geq 0$.

Here the dual problem is to maximize $w = 5y_1 + 8y_2 + 6y_3 + 7y_4 + 4y_5$ subject to the constraints

$$\begin{pmatrix} 1 & 5 & 1 & 2 & 1 \\ 2 & 3 & 2 & 1 & 1 \\ 1 & 2 & 2 & 1 & 1 \\ 3 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{pmatrix} \leq \begin{pmatrix} 2 \\ 3 \\ 2 \\ 3 \end{pmatrix}.$$

Adding in slack variables, these inequalities are equivalent to the system of equations whose augmented matrix is

$$\begin{pmatrix} 1 & 5 & 1 & 2 & 1 & 1 & 0 & 0 & 0 & 2 \\ 2 & 3 & 2 & 1 & 1 & 0 & 1 & 0 & 0 & 3 \\ 1 & 2 & 2 & 1 & 1 & 0 & 0 & 1 & 0 & 2 \\ 3 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 3 \end{pmatrix}$$

Now the obvious solution is feasible so there is no hunting for an initial obvious feasible solution required. Now add in the row and column for w. This yields

$$\begin{pmatrix} 1 & 5 & 1 & 2 & 1 & 1 & 0 & 0 & 0 & 0 & 2 \\ 2 & 3 & 2 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 3 \\ 1 & 2 & 2 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 2 \\ 3 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 3 \\ -5 & -8 & -6 & -7 & -4 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

It is a maximization problem so you want to eliminate the negatives in the bottom left row. Pick the column having the one which is most negative, the -8. The pivot is the top 5. Then apply the simplex algorithm to obtain

$$\begin{pmatrix} \frac{1}{5} & 1 & \frac{1}{5} & \frac{2}{5} & \frac{1}{5} & \frac{1}{5} & 0 & 0 & 0 & 0 & \frac{2}{5} \\ \frac{7}{5} & 0 & \frac{7}{5} & -\frac{1}{5} & \frac{2}{5} & -\frac{3}{5} & 1 & 0 & 0 & 0 & \frac{9}{5} \\ \frac{3}{5} & 0 & \frac{8}{5} & \frac{1}{5} & \frac{3}{5} & -\frac{2}{5} & 0 & 1 & 0 & 0 & \frac{6}{5} \\ \frac{14}{5} & 0 & \frac{4}{5} & \frac{3}{5} & \frac{4}{5} & -\frac{1}{5} & 0 & 0 & 1 & 0 & \frac{13}{5} \\ -\frac{17}{5} & 0 & -\frac{22}{5} & -\frac{19}{5} & -\frac{12}{5} & \frac{8}{5} & 0 & 0 & 0 & 1 & \frac{16}{5} \end{pmatrix}.$$

There are still negative entries in the bottom left row. Do the simplex algorithm to the column which has the $-\frac{22}{5}$. The pivot is the $\frac{8}{5}$. This yields

$$\begin{pmatrix} \frac{1}{8} & 1 & 0 & \frac{3}{8} & \frac{1}{8} & \frac{1}{4} & 0 & -\frac{1}{8} & 0 & 0 & \frac{1}{4} \\ \frac{7}{8} & 0 & 0 & -\frac{3}{8} & -\frac{1}{8} & -\frac{1}{4} & 1 & -\frac{7}{8} & 0 & 0 & \frac{3}{4} \\ \frac{3}{8} & 0 & 1 & \frac{1}{8} & \frac{3}{8} & -\frac{1}{4} & 0 & \frac{5}{8} & 0 & 0 & \frac{3}{4} \\ \frac{5}{2} & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & -\frac{1}{2} & 1 & 0 & 2 \\ -\frac{7}{4} & 0 & 0 & -\frac{13}{4} & -\frac{3}{4} & \frac{1}{2} & 0 & \frac{11}{4} & 0 & 1 & \frac{13}{2} \end{pmatrix}$$

and there are still negative numbers. Pick the column which has the $-13/4$. The pivot is the $3/8$ in the top. This yields

$$\begin{pmatrix} \frac{1}{3} & \frac{8}{3} & 0 & 1 & \frac{1}{3} & \frac{2}{3} & 0 & -\frac{1}{3} & 0 & 0 & \frac{2}{3} \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 1 \\ \frac{1}{3} & -\frac{1}{3} & 1 & 0 & \frac{1}{3} & -\frac{1}{3} & 0 & \frac{2}{3} & 0 & 0 & \frac{2}{3} \\ \frac{7}{3} & -\frac{4}{3} & 0 & 0 & \frac{1}{3} & -\frac{1}{3} & 0 & -\frac{1}{3} & 1 & 0 & \frac{5}{3} \\ -\frac{2}{3} & \frac{26}{3} & 0 & 0 & \frac{1}{3} & \frac{8}{3} & 0 & \frac{5}{3} & 0 & 1 & \frac{26}{3} \end{pmatrix}$$

which has only one negative entry on the bottom left. The pivot for this first column is the

$\frac{7}{3}$. The next tableau is

$$\begin{pmatrix} 0 & \frac{20}{7} & 0 & 1 & \frac{2}{7} & \frac{5}{7} & 0 & -\frac{2}{7} & -\frac{1}{7} & 0 & \frac{3}{7} \\ 0 & \frac{11}{7} & 0 & 0 & -\frac{1}{7} & \frac{1}{7} & 1 & -\frac{6}{7} & -\frac{3}{7} & 0 & \frac{2}{7} \\ 0 & -\frac{1}{7} & 1 & 0 & \frac{2}{7} & -\frac{2}{7} & 0 & \frac{5}{7} & -\frac{1}{7} & 0 & \frac{3}{7} \\ 1 & -\frac{4}{7} & 0 & 0 & \frac{1}{7} & -\frac{1}{7} & 0 & -\frac{1}{7} & \frac{3}{7} & 0 & \frac{5}{7} \\ 0 & \frac{58}{7} & 0 & 0 & \frac{3}{7} & \frac{18}{7} & 0 & \frac{11}{7} & \frac{2}{7} & 1 & \frac{64}{7} \end{pmatrix}$$

and all the entries in the left bottom row are nonnegative so the answer is $64/7$. This is the same as obtained before. So what values for $\mathbf{x}$ are needed? Here the basic variables are y_1, y_3, y_4, y_7. Consider the original augmented matrix, one step before the simplex tableau.

$$\begin{pmatrix} 1 & 5 & 1 & 2 & 1 & 1 & 0 & 0 & 0 & 0 & 2 \\ 2 & 3 & 2 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 3 \\ 1 & 2 & 2 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 2 \\ 3 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 3 \\ -5 & -8 & -6 & -7 & -4 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

Permute the columns to put the columns associated with these basic variables first. Thus

$$\begin{pmatrix} 1 & 1 & 2 & 0 & 5 & 1 & 1 & 0 & 0 & 0 & 2 \\ 2 & 2 & 1 & 1 & 3 & 1 & 0 & 0 & 0 & 0 & 3 \\ 1 & 2 & 1 & 0 & 2 & 1 & 0 & 1 & 0 & 0 & 2 \\ 3 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 3 \\ -5 & -6 & -7 & 0 & -8 & -4 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

The matrix B is

$$\begin{pmatrix} 1 & 1 & 2 & 0 \\ 2 & 2 & 1 & 1 \\ 1 & 2 & 1 & 0 \\ 3 & 1 & 1 & 0 \end{pmatrix}$$

and so B^{-T} equals

$$\begin{pmatrix} -\frac{1}{7} & -\frac{2}{7} & \frac{5}{7} & \frac{1}{7} \\ 0 & 0 & 0 & 1 \\ -\frac{1}{7} & \frac{5}{7} & -\frac{2}{7} & -\frac{6}{7} \\ \frac{3}{7} & -\frac{1}{7} & -\frac{1}{7} & -\frac{3}{7} \end{pmatrix}$$

Also $\mathbf{b}_B^T = \begin{pmatrix} 5 & 6 & 7 & 0 \end{pmatrix}$ and so from Corollary 6.5.3,

$$\mathbf{x} = \begin{pmatrix} -\frac{1}{7} & -\frac{2}{7} & \frac{5}{7} & \frac{1}{7} \\ 0 & 0 & 0 & 1 \\ -\frac{1}{7} & \frac{5}{7} & -\frac{2}{7} & -\frac{6}{7} \\ \frac{3}{7} & -\frac{1}{7} & -\frac{1}{7} & -\frac{3}{7} \end{pmatrix} \begin{pmatrix} 5 \\ 6 \\ 7 \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{18}{7} \\ 0 \\ \frac{11}{7} \\ \frac{2}{7} \end{pmatrix}$$

which agrees with the original way of doing the problem.

Two good books which give more discussion of linear programming are Strang [25] and Nobel and Daniels [20]. Also listed in these books are other references which may prove useful if you are interested in seeing more on these topics. There is a great deal more which can be said about linear programming.

6.6 Exercises

1. Maximize and minimize $z = x_1 - 2x_2 + x_3$ subject to the constraints $x_1 + x_2 + x_3 \leq 10$, $x_1 + x_2 + x_3 \geq 2$, and $x_1 + 2x_2 + x_3 \leq 7$ if possible. All variables are nonnegative.

2. Maximize and minimize the following if possible. All variables are nonnegative.

   (a) $z = x_1 - 2x_2$ subject to the constraints $x_1 + x_2 + x_3 \leq 10$, $x_1 + x_2 + x_3 \geq 1$, and $x_1 + 2x_2 + x_3 \leq 7$

   (b) $z = x_1 - 2x_2 - 3x_3$ subject to the constraints $x_1 + x_2 + x_3 \leq 8$, $x_1 + x_2 + 3x_3 \geq 1$, and $x_1 + x_2 + x_3 \leq 7$

   (c) $z = 2x_1 + x_2$ subject to the constraints $x_1 - x_2 + x_3 \leq 10$, $x_1 + x_2 + x_3 \geq 1$, and $x_1 + 2x_2 + x_3 \leq 7$

   (d) $z = x_1 + 2x_2$ subject to the constraints $x_1 - x_2 + x_3 \leq 10$, $x_1 + x_2 + x_3 \geq 1$, and $x_1 + 2x_2 + x_3 \leq 7$

3. Consider contradictory constraints, $x_1 + x_2 \geq 12$ and $x_1 + 2x_2 \leq 5, x_1 \geq 0, x_2 \geq 0$. You know these two contradict but show they contradict using the simplex algorithm.

4. Find a solution to the following inequalities for $x, y \geq 0$ if it is possible to do so. If it is not possible, prove it is not possible.

   (a) $$\begin{array}{c} 6x + 3y \geq 4 \\ 8x + 4y \leq 5 \end{array}$$

   (b) $$\begin{array}{c} 6x_1 + 4x_3 \leq 11 \\ 5x_1 + 4x_2 + 4x_3 \geq 8 \\ 6x_1 + 6x_2 + 5x_3 \leq 11 \end{array}$$

   (c) $$\begin{array}{c} 6x_1 + 4x_3 \leq 11 \\ 5x_1 + 4x_2 + 4x_3 \geq 9 \\ 6x_1 + 6x_2 + 5x_3 \leq 9 \end{array}$$

   (d) $$\begin{array}{c} x_1 - x_2 + x_3 \leq 2 \\ x_1 + 2x_2 \geq 4 \\ 3x_1 + 2x_3 \leq 7 \end{array}$$

(e) $$\begin{aligned} 5x_1 - 2x_2 + 4x_3 &\leq 1 \\ 6x_1 - 3x_2 + 5x_3 &\geq 2 \\ 5x_1 - 2x_2 + 4x_3 &\leq 5 \end{aligned}$$

5. Minimize $z = x_1 + x_2$ subject to $x_1 + x_2 \geq 2,\ x_1 + 3x_2 \leq 20,\ x_1 + x_2 \leq 18$. Change to a maximization problem and solve as follows: Let $y_i = M - x_i$. Formulate in terms of y_1, y_2.

To see Chapter 7-12 download
Linear Algebra II Spectral Theory and Abstract Vector Spaces

To see Chapter 13-15 download
Linear Algebra III Advanced topics